U.S. Office of Education

Guidance testing

Chicago

1948

Business

and Technical

U. S. *Office of Education. Division of Vocational Education.*

Guidance testing, by Clifford P. Froehlich, specialist for training guidance personnel and Arthur L. Benson, specialist, individual inventory and counseling techniques, Occupational Information and Guidance Service. Chicago, Science Research Associates, 1948.

viii, 104 p. 23 cm.

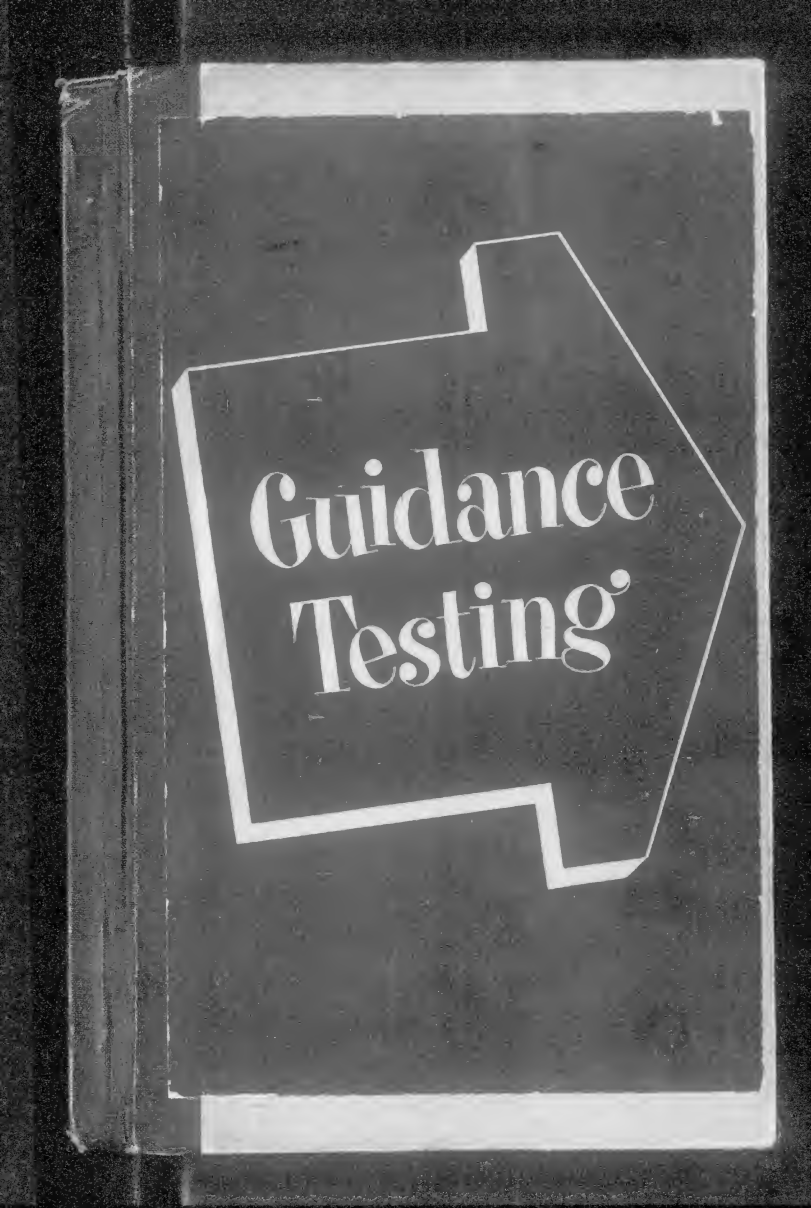
"Prepared by the Occupational Information and Guidance Service of the Division of Vocational Education in cooperation with the Division of Secondary Education, U. S. Office of Education."

"A basic library on testing": p. 88.

1. Personnel service in education. 2. Ability—Testing. I. Froehlich, Clifford Payo, 1914– II. Benson, Arthur L. III. Title.

LB1027.5.U57 371.26 48–1899 rev*

Library of Congress [r59d2]


Guidance
Testing

GUIDANCE TESTING

GUIDANCE TESTING

PREPARED by the Occupational Information and Guidance Service of the Division of Vocational Education in cooperation with the Division of Secondary Education, U. S. Office of Education, Federal Security Agency, Washington 25, D. C. The following persons acted as a committee of consultants who, by individual suggestions and in a conference of the whole, reviewed the manuscript in its successive steps:*

WALTER V. BINGHAM
Chief Psychologist
Adjutant General's Office
War Department

JOHN G. DARLEY
Professor of Psychology and
Director, Student Counseling Bureau
University of Minnesota

MITCHELL DREESE
Professor of Educational Psychology
George Washington University

HAROLD A. EDGERTON
Professor of Psychology and
Director, Occupational Opportunities Service
Ohio State University

DAVID SEGEL
Specialist, Tests and Measurements
Division of Secondary Education
U. S. Office of Education

ROBERT L. THORNDIKE
Associate Professor of Educational Psychology
Columbia University

ARTHUR E. TRAXLER
Associate Director
Educational Records Bureau

*The title given for each consultant indicates the position which he held at the time of consultation. Several of the consultants have recently assumed new responsibilities.

GUIDANCE TESTING

by

CLIFFORD P. FROEHLICH
Specialist for Training Guidance Personnel

and

ARTHUR L. BENSON
Specialist, Individual Inventory and Counseling Techniques

Occupational Information and Guidance Service
Division of Vocational Education
U.S. Office of Education
FEDERAL SECURITY AGENCY

SCIENCE RESEARCH ASSOCIATES, CHICAGO, 1948

FOREWORD

THIS BOOK IS ADDRESSED to those individuals who are faced with the responsibility of carrying on a guidance program in which they must directly or indirectly administer and interpret tests, even though their training in tests and measurements is limited. It deals with questions they must answer, such as: What is the place of testing in the guidance program? What things must be considered in planning a testing program? How are the tests selected? What should be measured? How are tests results used in the guidance program? A chapter is devoted to answering each of these questions.

The many different and sometimes contradictory approaches and concepts in the field of psychological measurement are confusing to the beginning counselor. The U. S. Office of Education believed that a summary statement on testing for guidance purposes, based on the pooled judgment of persons with extensive experience in testing, was needed. Accordingly, the persons listed opposite the title page were asked to serve as consultants to the Office for the preparation of this book. After a preliminary outline was submitted to them, these consultants met in Washington to discuss the content in detail. A stenographic transcript of that conference was supplied as a guide to the writers. The book was written on the foundation of this preparatory work, and then reviewed by each of the consultants. The U. S. Office of Education is appreciative of the services of these authorities who willingly gave of their valuable time to this work.

In addition to the consultants, many others have cooperated in the preparation of this book. The services of David Segel of the Division of Secondary Education were made available by Galen Jones, Director of that Division in the U. S. Office of Education. Francis G. Cornell, formerly Chief, Research and Statistical Service, reviewed the sections dealing with statistics and assisted in the preparation of Appendix B. State Supervisors of Occupational Information and Guidance assisted by reviewing the outline at their Seventh National Conference and by individual review of the final draft.

This book has been prepared by Clifford P. Froehlich and Arthur L. Benson, under the direction of Harry A. Jager, Chief of the Occupational Information and Guidance Service.

The Occupational Information and Guidance Service produced this manuscript for publication by the Government Printing Office. Printing funds were, however, unavailable and the Service was faced with the fact that not only was there a pressing need expressed throughout the field for the material, but also that its value was dependent to a large degree on timeliness. Therefore, permission was secured for obtaining proposals from private publishers. It is a pleasure under these circumstances to have enlisted the cooperation of Science Research Associates as the publishing agents for the manuscript.

RAYMOND W. GREGORY
Assistant U. S. Commissioner
for Vocational Education
U. S. Office of Education

Washington, D.C.
February, 1948

CONTENTS

Chapter 1

PLACE OF TESTING IN THE GUIDANCE PROGRAM

MEMBERS OF THE school staff are engaged in a wide variety of activities which taken collectively are called the guidance program. These activities include helping Jane find and advance in a job suitable for her; assisting Raymond in deciding whether he should select the academic or vocational curriculum; recommending that Helen's training in reading be deferred another six months; and helping John untangle his personality problems.

In each of these activities, we need to know a great deal about these four pupils. Does Jane's school history and record of class work indicate that she is trained for the job in which she is interested? Does Raymond's home background suggest that his parents will be able to help him attend college? Do the results of Helen's reading-readiness test predict failure for her if she is immediately assigned to a beginning reading class? Will taking part in co-curricular activities help John solve his problem?

These questions indicate only a few areas of information needed to assist these pupils in attacking their problems. Besides school history and record of class work, home background, and special aptitudes, we may list at least seven other types of information to which we may refer: mental ability or academic aptitude, achievement and growth in different fields of study, health, out-of-school experiences, educational and vocational interests, personality, and plans for the future.[1]

INFORMATION NEEDED FOR COUNSELING

If the guidance program is a well-developed one, a great deal of data in the areas already mentioned is available in each pupil's cumulative record. How is this information accumulated? We probably have interviews with Jane and Raymond several times early in their school careers. During these interviews, both pupils give facts or express attitudes which are recorded in their cumulative records. Both pupils fill out questionnaires at various times. Some of their teachers make anecdotal records or behavior descriptions based on their observations of these pupils in class or on the playground. Both pupils may have an opportunity in their English classes to write autobiographical themes. Teachers rate them on such traits as

[1]Arthur E. Traxler, *Techniques of Guidance* (New York: Harper & Brothers, 1945), pp. 20-25.

initiative, cooperativeness, or sociability. Information regarding attendance, subjects taken, and marks is transcribed from their administrative records. Periodically, they are given some kind of a physical examination. And, finally, they take some tests. We shall be concerned, then, in this book with only one of many sources of information about pupils; namely, testing. By means of it we can obtain information which enables us to help the pupil with his problems.

DIFFERENT TECHNIQUES GIVE SAME TYPE OF INFORMATION

Several different techniques might be used to get the same type of information. Jane's vocational interest can be estimated by an interview, by her responses on a questionnaire, or by her scores on an interest test. Raymond's mechanical aptitude can be revealed by anecdotal records describing unusual projects he has completed out of school, by his autobiographical theme discussing his favorite hobby, or by his scores on a mechanical aptitude test.

Is this unnecessary duplication? If we establish a comprehensive testing program in our school, can we dispense with some of the other techniques for collecting information? We probably could if the same information from all sources led to the same conclusions. In the interest of economy, we should probably abandon our testing program and depend upon our other less expensive sources. Experience, however, does not reveal a high correlation of results. Jane's statement regarding her interests is frequently at variance with the results of her interest test. Raymond's high mechanical aptitude inferred from his anecdotal records is not always confirmed by his mechanical aptitude test score. Frequently, test results and data gathered by other means appear contradictory.

EVALUATE INFORMATION BEFORE MAKING DECISIONS

We can think of at least two reasons for these discrepancies. First, our data may be erroneous. Raymond's father, a skilled craftsman, may be responsible for the fine projects we have noted on Raymond's anecdotal records. There may be an error in adding the sub-scores on the different parts of his mechanical aptitude test. Our information may simply be false.

Second, our interpretation of the data may be incorrect. Jane's excellent school record in subjects closely related to her job interest may not mean that she is adequately trained for the job. Perhaps we have interpreted her school record as a pure measure of her achievement in this training. Actually, it may not be a pure measure of her accomplishments because it is weighted heavily with the personal friendship which exists between Jane and her teacher. Our information may not mean what we think it means.

Once we recognize that all data are subject to these two kinds of error, we can see several reasons for using all the sources of information at our

command. The degree of confidence we feel in any single item is considerably increased if we can draw from other sources of information substantially the same conclusions. Thus test results which merely confirm conclusions based on interviews, ratings, and the like serve an important function.

USE ALL SOURCES OF INFORMATION

Another reason for using all methods for information-gathering is suggested by the frequent discrepancies in the data obtained from different sources. We have seen how such inconsistencies can help us discover errors in fact and errors in interpretation. We shall see in Chapter V how comparison of these differences can be used as a basis for discovering other facts about pupils which would not be apparent if we limited our sources of information. Our testing program should supplement, not be a substitute for other sources of information in the individual inventory.

TESTS SUPPLEMENT OTHER DATA

Regardless of guidance activities, Jane will probably find a job. Raymond will select a curriculum, and Helen will be assigned to a class. Life will go on; adjustments will be made independently. We can probably help these young people make more intelligent decisions or better adjustments, however, even if we counsel them only on the basis of non-test data. We shall not do a perfect job even if we do include test scores in our individual inventory. But there is ample evidence that we can do a better job if we add test results to other relevant data.

One other aspect of our testing program is suggested by the illustrations we have used. We have considered testing as a technique for helping a single individual meet a specific need. Unless we can state precisely how we will use the results of a test so as to help Jane, or Raymond, or Helen to solve one of their problems, we shall have difficulty in justifying the administration of that test.

Fortunately, the problems of individual pupils are frequently the problems of groups. Whether or not Helen is ready to learn to read is a question we can ask regarding most of the other first-graders. It is clear that we can test groups of individuals most economically under such circumstances.

TESTING SHOULD MEET INDIVIDUAL NEEDS

On the other hand, if Jane is seeking a part-time job so that she can continue in school, she may be the only pupil with this particular problem at the time. Jane cannot wait to make her decision, and we do not want to deprive her of a service we give to her more typical classmates. Another pupil may be confronted with an unusual problem in the solution of which some test not included in the regular testing program would be helpful. We must be alert to the needs for information

which are not met by a general testing plan formulated on the typical problems of pupils.

The testing program in any school may be thought of as having three aspects. The first concerns the tests given to shed light on administrative or instructional problems. Usually these tests are given to large blocks of pupils. The results of this type of testing are frequently useful to the guidance program as it deals with each pupil individually. Consequently, guidance workers should have a part in planning this program.

A second aspect of the school's testing program deals with group tests given for guidance purposes. In counseling we may find that a large proportion of pupils need the results of an interest test. It may be more economical to give this interest test to all pupils as a group rather than to selected ones individually. Economy is the only justification for group testing for guidance purposes. It is wise to coordinate this testing with other group testing in the school.

The third phase of testing is also carried on within the guidance program. It is concerned with the administration of tests to individuals to meet their needs. Unification of these three aspects of testing into an overall program is essential. Plans for testing at all levels in the school and for all purposes should be coordinated.

It is equally important that a uniform system for recording all standardized test data be adopted. Too frequently tests are given to throw light on some particular administrative problem without any provisions being made for utilizing the individual data for guidance purposes. If a test is worth giving, the results should be recorded. And if they are worth recording, the records should be available to those who have use for them.

PROFESSIONAL LEADERSHIP IS NECESSARY FOR TESTING

In the small school, the counselor will probably be in the most favorable position to assume professional leadership of the testing program. We have accepted the problems that Jane, Raymond, Helen, and John face, as our problems. We must also prepare ourselves to accept the responsibility for leadership in organizing and administering a testing program which fits the needs of our school.

Chapter II

PLANNING A TESTING PROGRAM

WHILE WE HAVE recognized that testing should meet the needs of individuals, it is obvious that all pupils have certain basic needs in common. It is in this area that a general testing program can be developed. At the same time it should be remembered that individual problems will frequently indicate the desirability of additional test information not generally needed at any given time.

FOUR BASIC CONSIDERATIONS IN PLANNING THE PROGRAM

In planning the testing program, there are at least four basic considerations. Although there is considerable overlapping among the specific applications of these concepts, it would seem profitable to discuss each one separately.

First, the testing program should be a cooperative enterprise on the part of teachers, pupils, and parents. The entire program should be based first, on the results of a study by the school staff of the need for test information in dealing with pupils. This study should include consideration of the use of test results in attacking instructional and administrative as well as guidance problems. It may well comprise plans for evaluating different methods of teaching.

COOPERATIVE PLANNING IS ESSENTIAL

One warning note is necessary here. The guidance testing program should avoid the administrative problem of evaluating teachers. Such studies often lead to erroneous conclusions. For example, the fact that Miss Hawkin's classes have a lower average score on an achievement test than Mrs. Foreman's is no indication of the relative effectiveness of these two teachers. Such a difference may be found because (1) pupils differ in ability to learn what is being taught; (2) the test does not measure what one or both of these teachers are trying to teach; or (3) the test is somewhat unreliable. Evaluation of teachers on the basis of their pupils' performance on achievement tests is a highly technical process. It should be attempted only by persons with special training in research testing. To include the comparison of teachers as one of the aims of the testing program invites the antagonism of teachers who have learned by bitter experience the unjustified conclusions which are frequently made on the basis of unsophisticated studies in this field.

The cooperation of pupils also in carrying out the program is important. Pupils should understand the purposes for which tests are given so as to effect adequate motivation without unnecessary tenseness. They should also know that the test results will be interpreted to them.

We should enlist the cooperation of parents in the program. Discussions with them of the school's need to use information revealed by tests and interpretation of test results can be mutually profitable. In Chapter VI, the thesis is developed that pupils should be told as much about their test performance as they can correctly interpret and are ready and able to act upon. This rule is an equally valid basis for determining what test information should be given to parents. As a minimum, we can keep them informed of the kind of test information we are accumulating, the types of problems on which test information can throw some light, and how the results are used in our guidance activities. With some groups of parents, the discussions can go so far as to include simple explanations of concepts basic to interpretation of test results. Test data for a particular child can be discussed, of course, only in an interview. There, the willingness and ability of the parent to take constructive action on any particular information can be estimated. Our judgment on this point will determine the type and extent of information we can give. Of one thing we can be sure: Misinformed parents can kill the testing program; enlightened parents can assist us in doing a better job of helping their children.

Second, the testing program should be a long-range program. It must be conceived as a continuing project for collecting information about each individual as the need for such information arises.

LONG-RANGE PLANNING NECESSARY

Such a program will envisage the gathering over a period of years of test evidence for each pupil. It is apparent that the recording of such data must be systematic and complete to be useful over a long period. As changes take place in the educational or vocational environment, the needs of the pupils will change. Likewise, new and better tests may be constructed or more may be learned about tests available but not used now. The testing program, therefore, must be adaptable to change.

Third, the testing program should be practicable. What is practicable for one school may be out of the question for another, but there are two general rules that can be helpful in most situations:

PROGRAM MUST BE PRACTICABLE

(1) The routine clerical or statistical work involved in scoring tests and recording results should be kept as low as possible to get the needed information. (2) The loss of time in the regular school schedule must not be out of proportion to the expected gains in instructional and counseling efficiency.

Another hurdle which the testing program must leap is cost. Comparison of test catalogs and price lists will demonstrate that there are considerable differences in prices of tests which apparently yield the same type of information. The costs of tests are not necessarily related to their usefulness. Many tests have been adapted for use with separate answer sheets so that only a few test booklets need be purchased. In the course of a few years, such tests may involve a smaller total outlay than tests not adapted for separate answer sheets, the initial cost of which may be considerably less. Some publishers rent test materials. Schools in a few areas have combined orders to increase purchasing power and reduce unit costs. In some states sponsoring a testing program, the State Supervisor of Occupational Information and Guidance makes arrangements for schools to participate. Even when no state testing program exists, this Supervisor frequently is able to help schools find solutions to testing problems. Sometimes it is possible to save money by buying combinations of tests from publishers and testing bureaus sponsoring national testing programs.

Fourth, the testing program should be professional. The value of any item in the individual inventory is dependent on our ability to use it constructively in helping a child adjust himself to his opportunities. If we have had little experience with standardized tests, we should start with a modest program involving one or two types of test. By so doing, we put ourselves in a better position to learn how to use the scores on each test in order to throw light on several different problems. We do not get so busy giving and scoring tests and recording scores that we can never find time to use the results. And we avoid the unfavorable reactions of parents, children, and teachers which sudden emphasis on testing may arouse.

PROFESSIONAL TRAINING BASIC TO EFFECTIVE OPERATION

Test scores, particularly those labeled "intelligence" or "personality," have a peculiar fascination for some people. If the testing program is to be professional, it must make provision for minimizing gossip. Frequently an in-service training program on the meaning of test scores will nip such activities in the bud. Again, the comparison of individuals is usually more invidious than helpful. How can Oscar or Henry be helped by the observation that Oscar's score was 61 while Henry's was only 23? The comparisons a counselor will find most helpful will be those in which Oscar is compared with himself. That Oscar gets very high scores on achievement tests, but average or low marks from his teachers, is an important discovery. It is a starting point for an investigation which can be helpful both to Oscar and his teachers.

If the testing program is directed toward helping Oscar, his teachers will not feel that low scores discredit their ability. They will not make

special preparation in their classes for the test. Oscar's social studies teacher should realize that she may do him a disservice if she gets copies of the social studies achievement test in advance, and by using it as a basis for her teaching helps him get a higher score. Oscar's achievement in social studies will no longer be comparable to his achievement in other fields. We have to help him make decisions often enough on data which are unavoidably subjective without basing those decisions on test data which have been influenced by coaching.

Finally, the testing program cannot be a professional one unless we understand the contribution which test scores may make to the guidance program. Henry may inform us, for example, that he is interested in becoming a surveyor, yet his scores on an interest test may be low in mathematics and science, and high in agricultural and mechanical areas. Investigation of this discrepancy may reveal that Henry filled out his questionnaire while a survey for a new highway through his father's farm was being made. We should probably be able to discover the transitory nature of his stated interest even if Henry had not taken an interest test. But without the test results, we should have no clue as to Henry's interest. The test score has helped us in two ways: First, it has led us to question the accuracy of other data in the individual inventory, and second, it has provided a starting point for the interview.

Henry's case may be carried one step further. As one result of this interview, Henry may decide that it would be worth while for him to take the interest test again. This time he continues to show low mathematical interests and high agricultural interests, but his score in science now is much higher, and in mechanical activities much lower than on the earlier test. We may conclude that the test is not as indicative as it might be. But, is Henry's statement of interest any more indicative? A professional attitude toward test scores requires us to recognize that all tests are not infallible indicators. But in recognizing this point, we must keep in mind the limitations of other sources of information. We probably can do a fair job of counseling Henry without any test scores, but we can do a better job if we have relevant test information.

CRITERIA FOR SELECTING KINDS OF TESTS

Now that we have outlined some of the requirements for a testing program, let us set up criteria which will help us to decide what types of tests we shall include in our program, and to decide when the tests should be administered so as to be most useful.

From the viewpoint of economy, we should select those types of tests which will yield information that will be valid for counseling pupils with regard to as many of their problems as possible.

SELECT TESTS RELEVANT TO MOST PROBLEMS

Other things being equal, a pupil's score on a test of general reading ability will be of value in counseling him regarding more of his problems than his score in an elementary algebra test. On this basis, we decide that general reading tests have priority over specific subject achievement tests in planning our program. The latter are of value, but their value is relatively less. We shall have to decide how extensive our program can be and then select tests with widest application.

We are also concerned with the immediate usefulness of the tests. When we can give only a limited number of tests, we should select those tests which provide information of immediate value. This does not mean that we shall not be able to develop long-range plans for testing. Although many of the tests may be useful at once it can be expected that some will be useful at a later date. A balance between immediate and delayed benefits should be our goal.

USE TESTS YIELDING DATA IMMEDIATELY USEFUL

Catherine may think that she wants to prepare for commercial art. We recognize that general scholastic aptitude tests and academic achievement tests are not particularly valid predictors of total achievement in art. They certainly do not give a basis for making judgments regarding her chances for success. For example, we may wonder about Catherine's sense of color because the color combination in her clothes is frequently poor. This may be the result of poor color vision or one of a dozen factors. We decide that Catherine's score on a test of color vision will be helpful. We should select the types of tests which supply information in those areas in which our available data are least relevant.

SELECT TESTS WHICH SUPPLY MISSING INFORMATION

Not only must we decide what types of tests to include in our program, but we must also determine when to test. The counselor and school staff can be of most assistance to a pupil by learning as much as possible about him as soon as they become responsible for his development. Obviously school entrance is a crucial time for gathering information. How extensive the process should be depends in part on how much meaningful information about the pupil is already available.

SCHOOL ENTRANCE GOOD TIME TO TEST

Additional testing becomes advisable also when we have reason to believe that the results of earlier tests are questionable. If Dorothy's score on a scholastic aptitude test is definitely out of line with her teacher's marks and other evidence of scholastic achievement, we may decide to give her a similar aptitude test to verify the results of the first test. Or we may adopt an hypothesis to explain this discrepancy and subject it to verification by giving her a different kind of test.

RETEST IF EARLIER RESULTS ARE QUESTIONABLE

Testing which has significance for choices, as we have already suggested, should be done when pupils need to make the choices. A pupil's vocational interest scores are of little use to him if he has no opportunity to make choices on the basis of his scores.

When achievement tests have been used as an important basis for recommending corrective or remedial action, the periodic repetition of similar tests for the purpose of evaluating these recommendations is desirable. We should make an effort to discover whether or not the program Carl has undertaken to correct his reading deficiency is achieving the desired result. Does he need more of the same kind of training? Is a different approach to his problem indicated? Or has he reached a standard which is appropriate for him?

TESTS USEFUL IN EVALUATING REMEDIAL ACTION

Wide differences in school administration and organization in kinds of non-test information in the individual inventory, in occupational and educational opportunities, and in numerous other factors make detailed description of the types of tests which are most useful at various grade levels impossible. In a small school, the odds are that we shall not be able to do extensive testing at each grade level every year. We are, however, able to select a few grades in which test information can be most helpful and make general use of tests which supply the information needed at those grade levels. We shall remember, of course, that we are not really testing a *grade* or a *class*, but *individuals*, so that the individuals tested in any grade-level scheme will often include many from adjacent grades.

PLAN PROGRAM TO FIT LOCAL NEEDS

Chapter III

DECIDING WHAT TO MEASURE WITH TESTS

In the previous chapter we observed that any data in the individual inventory are valid to the extent that they really mean what we think they mean. And we noted that such data are reliable to the extent that in repeated instances we obtain consistently similar information. Now let us consider the reliability and validity of test data. Fairly precise methods of determining these characteristics of tests have been devised. One frequently used method is the computation of correlation coefficients.

If Edgar gets the highest score in his class today on an arithmetic test, other things being equal, we can expect him to get the highest score in a similar test tomorrow. Suppose we list five members of his class in order of their scores. Each pupil ranked in the same order on the second test as he did on the first. Each pupil's rank on the first test predicts perfectly what his rank on the second test will be. The relation between the two tests is perfect. Statisticians, who have worked out mathematical formulae for expressing this relationship between two sets of data, would say that the coefficient of correlation is +1.00.

CORRELATION COEFFICIENTS EXPRESS RELATIONSHIP NUMERICALLY

Pupil	Rank on first test	Rank on second test
Edgar	1	1
Joan	2	2
Paul	3	3
Tom	4	4
Frieda	5	5

But if the situation is exactly reversed, as in the second illustration, and each pupil who scores high on the first test scores low on the second, the coefficient of correlation is —1.00.

Pupil	Rank on first test	Rank on second test
Edgar	1	5
Joan	2	4
Paul	3	3
Tom	4	2
Frieda	5	1

In both cases accurate predictions of rank on the second test can be made if we know the rank on the first. If the positive sign is used with the correlation coefficient, however, we know that good performance on the first test is positively related to good performance on the second test. But if the negative sign is used, we expect high scores on the first test to be related to low scores on the second.

The correlation coefficient thus ranges from +1.00 through 0.00, where no relationship exists between the two sets of data, to —1.00. Perfect correlation coefficients are rare. Negative correlations are not very common either. Contrary to frequently expressed notions, it is not true that pupils who are very good in one activity are usually very poor in something else. They may be only mediocre in the second undertaking so the correlation between the two abilities may be low, but it is usually positive.

Ordinarily, our data for the results on the two arithmetic tests can be expected to be similar to that presented in the third illustration. Joan does better than Edgar, and Tom does better than Paul on the second test, while Frieda remains the lowest of our five. The correlation coefficient computed by rank difference method for these data is +.80.

Pupil	Rank on first test	Rank on second test
Edgar	1	2
Joan	2	1
Paul	3	4
Tom	4	3
Frieda	5	5

HOW CONSISTENTLY DOES THE TEST MEASURE?

The correlation coefficient then indicates the degree of relationship between two sets of data. When this method is used to determine the consistency with which tests measure whatever they measure, test makers usually call it the *reliability coefficient.* Thus in the above example we can conclude that the test is fairly reliable because the results of the second test are moderately consistent with the results of the first.

Test manuals indicate several different methods of obtaining the two sets of data necessary to compute the reliability coefficient. The method described previously is commonly called the test-retest method. When only one test is available a common practice is to score each pupil's test by first counting only the odd-numbered items in the test, and a second time counting only the even-numbered items. Since the correlation coefficient of these two sets of scores is a measure of the reliability of two tests half as long as the original test, this coefficient is usually corrected or stepped up by applying a formula which gives an estimate of the reliability if each of our half-tests had been twice as long. The coefficient so obtained is called the corrected odd-even, corrected split-half, or Spearman-Brown reliability coefficient.

METHOD OF ESTIMATING AFFECTS SIZE OF RELIABILITY COEFFICIENTS

There are other methods of computing the reliability coefficients of tests, but these two are the most common. These two methods yield slightly different estimates of the reliability of any one test. Test makers have found a good many factors which make minor differences in the reliability coefficient of a test, so we need not attach too much importance to differences of only .03 or .04.

Reliability coefficients have one serious limitation. If there are large differences in arithmetic achievement among the pupils in our group, the rank of each pupil will be about the same on each test. But if the differences in achievement among our pupils are very slight, the two tests will not agree nearly so well. For tests of ability and achievement, the range of ability of persons in the group has an important effect on the size of the coefficient. We must be somewhat cautious in accepting a reliability coefficient at its face value if the group from which it was computed is not described.

We shall have to rely on our experience and our knowledge of the reliability of available tests in order to set up some standards for selecting tests. In general, if we wish to use the results with individuals, we should select tests which have reliability coefficients of .85 or better. In selecting achievement or ability tests, we may require that the reliability coefficient be at least .85 for pupils at the same grade level as the group we wish to test; if two or three grades were combined in computing the reliability, we may demand that the coefficients be .90; and if four or five grades were used, about .95.

DESIRABLE RELIABILITY COEFFICIENTS

Some tests are scored by parts so that part-scores as well as total scores are obtained. If we hope to use Paul's four scores in addition, subtraction, multiplication, and division as well as his total arithmetic score, each of these part-scores must meet high standards of reliability. The reliability needs to be high when we are dealing with differences between highly

correlated scores. The reliability need not be so high when the correlation between scores is low.

WHAT DOES THE TEST MEASURE?

Even after we have concluded that a test reliably measures whatever it is intended to measure, we still have the problem of deciding just what a high or low score on the test means. The crucial problem in guidance testing is validity. Frank's score on a mechanical aptitude test may be highly reliable. His score on the odd-numbered items of the test may be identical to his score on the even-numbered items. His score on one form of the test may be the same as his score a week later on a second form of the test. But until we know what this score means, we are at a loss to interpret our data.

One clue we have is the title of the test—Mechanical Aptitude. This suggests that the pupil's high score on the test means that he can learn quickly and easily the skills required in some of the shop courses offered in our school. Suppose that we advise him to try some of these courses. He does so. He finds them very difficult and gets low marks. The shop teacher complains that while the pupil seems interested in doing his work, he is so clumsy that he botches most of his projects. Clearly his test score does not mean what we thought it did. On reviewing the test, we may find that it did not require the pupil to demonstrate his motor coordination or manual dexterity, although it did require him to solve problems involving mechanical comprehension and to have a wide range of information about tools. *The title of a test does not always tell us what the test measures.*

TEST TITLES ARE NOT ALWAYS MEANINGFUL

Test makers frequently attack the problem of the meaning of the test scores by obtaining the coefficient of correlation between the test scores of a group and some other measure of performance of the same group. This second measure becomes the standard, or criterion, by which the first is judged. To compute a correlation coefficient always requires having two sets of data for a single group, one of which is the test scores. In a scholastic aptitude test, the other set of data may be the average of teachers' marks for each pupil during the following year. It is the criterion we accept as a measure of scholastic success. A correlation of .60 between these two sets of data shows that the test scores predict pupils' marks fairly well.

CRITERION IS NEEDED FOR VALIDITY

We do not find validity coefficients quoted in test manuals as frequently as reliability coefficients. Test-makers have considerable difficulty finding a suitable criterion with which to match their test scores. With our scholastic aptitude test, we had to wait a year after the test had been given before the average of teachers' marks was available. Even then, the criterion itself

was probably not very reliable. Perhaps we should have waited longer and averaged marks for two or three years. The reliability of the criterion is frequently lower than that of the test. Thus the coefficients of correlation between the test scores and the criterion are limited by the unreliabiliy of both the test and the criterion, itself. For this reason, we cannot expect to find very high validity coefficients. If the criterion is itself reliable and if the test is actually a valid measure of the ability represented by the criterion, we would expect that the validity coefficient would approach the reliability coefficient for the test. Validity coefficients below .30 indicate a negligible correlation between the test and the criterion. We seldom find validity coefficients over .70. If we do, we should consider the possibility that the test and the criterion both measure something other than what was intended to be measured.

One other device frequently used by test-makers to indicate the meaning of test scores is to measure the ability of the test to distinguish between groups known to be different. A personality test score has more meaning if we know that there is a statistically significant difference between the average score made by children in a mental hospital and the average score made by pupils in school. Nevertheless, we must be cautious in interpreting scores of tests which have been validated by this method. It would probably be easy to construct a test which would differentiate between pupils who had completed first-year algebra and those who had never studied algebra. On the basis of the test results we might be able to separate a group of unknown pupils into two groups, those who had studied and those who had not studied algebra, with 100 per cent accuracy. The test would be a highly valid instrument for this differentiation. But if we tried to use it as an achievement test, we might make some serious errors. The pupil who makes the highest score on such a test may not be the best algebra student. He simply knows more algebra than the non-algebra pupil. The odds are that he is not the poorest algebra student, but a test which has been devised to differentiate between fairly widely separated groups is not necessarily a good instrument to evaluate performance within one of those groups. When this method of validation has been used, we usually need concern ourselves with only those pupils whose scores are extremely high or low.

For many tests, however, we cannot find any statistical evidence of validity. In attempting to validate interest tests, the authors would like to give their test to, say, a large group of mechanics who were interested in their work. The scores of these mechanics could then be compared with the scores made by non-mechanics or mechanics who disliked their work. If there were important differences between the scores of these two groups, we could con-

NON-STATISTICAL EVIDENCE OF VALIDITY USEFUL

clude that the test had some validity as a measure of the special interests of mechanics. The practical difficulties encountered by the test-maker in getting a large number of scores from a single occupational group are very real. If, in addition, he attempts to separate the satisfied from the dissatisfied members of that occupation, he may so reduce the size of his groups that we will not have much confidence in the differences he finds.

Achievement tests at the elementary level are validated usually by noting the increase in the percent of correct answers from one grade to the next higher grade. As a rule, test-makers seldom attempt to validate achievement tests statistically. They construct a test which measures what is taught in School A. If the teachers in this school appraise the achievement of their pupils accurately the correlation coefficient between the test scores and teacher's marks indicates the validity of the test.

We find that School B has different objectives from School A and that the subject matter taught in School B was poorly covered by our test. From this school's point of view, the validity coefficient was a fraud. The test-makers' logic is good. With few exceptions, the judgment of subject-matter teachers on the validity of achievement tests covering what they teach is better than any available statistics.

The method used to determine the content of the test may help us determine what the test measures. We shall have to use our own judgment to some extent. But we should certainly consider the reputation of the test authors and publishers. Books listed in Appendix A include selected lists of tests which are generally considered to be valid measures of important characteristics of pupils. State supervisors of guidance can be asked to draw upon their experience with testing to assist in selecting tests. Particularly with achievement tests, interested faculty members should be satisfied that the content of the test is appropriate.

Some faculty members desire to construct their own achievement tests. They believe that these tailor-made tests give a better estimate of pupils' achievement than tests available commercially. They construct their tests to measure achievement in terms of the content and objectives of their teaching. The counselor not only should encourage these teachers, but he should also provide professional assistance. The construction of test questions which are both valid and reliable is difficult. The analysis of items requires much statistical computation. The revision or replacement of faulty items may consume many hours. If the result is a reliable and valid achievement test designed to meet the needs of the pupils, the expense in time and money is justifiable. But if the test is hastily constructed without regard for difficulty, ambiguity, reliability, or other factors, more satisfactory results can be obtained by using a standardized test. The standardized test may not dovetail with the curriculum perfectly, but presumedly all pupils have this same handicap, thus making their scores comparable.

HOW DO NORMS HELP US INTERPRET TEST SCORES?

Thus far, we have considered each pupil's performance on a test in terms of his raw score or of his rank in the class. This procedure becomes extremely unhandy when we compare his performance on several tests. His score of 58 on one test may be below the average of his class, while his score of 23 on another test may be well above the average. Madeline's rank of fifteenth in an English class of thirty-five pupils is certainly not the same as her rank of fifteenth in a French class of eighteen. Raw scores and ranks are hard to interpret.

NECESSITY OF NORM SCORES

Test-makers have tried to solve this difficulty by giving tables in their test manuals which enable us to convert raw scores into some kind of norms. The norms most commonly found are percentile ranks, age scores, and grade scores and standardized scores.[1] Madeline's standing in English and French can be easily changed to percentile ranks. Madeline ranks fifteenth in a class of thirty-five. The fourteen pupils who do better than she does constitute 40 percent of the class. Thus we can say that Madeline does as well or better than 60 percent of the class, or that her percentile rank is 60. In French, fourteen out of eighteen pupils, or 78 percent do better than she does. Hence her percentile rank, which is found by subtracting 78 from 100, is found to be 22. Percentile ranks give us a fair picture of her relative standing in English and French.

Suppose that together with our data about Madeline, we consider the scores of Mable and Harry, who rank fourteenth and sixteenth, respectively in the English class. Harry, who scores only 1 point less than Madeline, is 3 points lower in percentile rank, while Mabel, who scores 10 points more, is still only 3 points higher in percentile rank.

LIMITATIONS OF PERCENTILE RANKS

Pupil	English score	Rank in class	Percentile rank
Mabel	47	14	63
Madeline	37	15	60
Harry	36	16	57

If the test scores are correct in indicating that Mabel's performance is quite superior to Madeline's and that Madeline's and Harry's are about the same, the percentile ranks certainly obscure these facts. Of course, when

[1]The term *standardized score* is used throughout this book to designate scores based on standard deviation units of the normal curve. In the illustrations and in Appendix B the standardized scores used are a modification of T-scores which were originally proposed by McCall. In these cases one-tenth of a standard deviation is assigned to each unit on the scale and the mean is abitrarily placed at 50.

the percentile ranks are based on a larger sample, such errors are less likely to occur. With these advantages and limitations of percentile ranks in mind, let us look at that other commonly used norm, the standardized score.

USING STANDARDIZED SCORES

We can describe standardized scores without going into details of the statistics involved in computing them. Like percentile ranks, standardized scores ordinarily run from 0 to 100 with average performance indicated by 50. Here the similarity ceases. In percentile ranks the highest score on a test automatically becomes 100 since 100 percent of the scores are equal to or below that score. Standardized scores of 100 are very rare. In fact, standardized scores are so calculated that nearly 70 percent of all scores in the group on whom the test was standardized have scores between 40 and 60. Standardized scores are based on the actual raw score of each individual rather than his rank in the group.

Pupil	English Score	Rank in Class	Percentile Rank	Standardized Score
Mabel	47	14	63	56
Madeline	37	15	60	52
Harry	36	16	57	52

We can compare the three pupils once more. Mabel's and Harry's English scores show that Mabel's performance on the test was superior to Madeline's and Harry's, but give no indication how well any one of the pupils performed in relation to his class or any larger group. The ranks indicate the same facts, but we need to know that there were thirty-five pupils in the group in order to interpret these ranks. The percentile rank is the first of these scores which tells us how Madeline stands not only in relation to the other two pupils, but also in relation to the class as a whole. Both rank in class and percentile ranks, however, obscure the large difference between Madeline's and Mabel's performance and the small difference between Madeline's and Harry's performance. The standardized scores give us a fairly good idea of how these three pupils stand in relation to the rest of the class and also reflect more accurately the difference in performance between these pupils.

Thus far we have discussed percentile ranks and standardized scores which have been computed using the scores of a single class in our school.

LIMITATIONS OF NORMS

Norms for published tests are usually computed on test results from hundreds of students in several schools. These so-called national norms need to be interpreted with some caution. If we find that Frank's percentile rank on a geography test is 46, we know that Frank did as well as or better than 46 percent of

the group used in establishing the norms. To interpret Frank's performance we need to know something about this group. Have they had approximately the same amount of instruction in geography that Frank had? Obviously if the test norms were established on a group of eighth-grade pupils, then fifth-grade Frank's performance below the fiftieth percentile does not have the same significance it would if Frank were an eighth-grader.

We shall have even greater difficulty if we try to compare Frank's percentile rank of 46 in geography with his percentile rank of 58 in arithmetic. Unless both of these tests were standardized on groups having an educational background similar to Frank's, we cannot say that Frank's achievement in arithmetic is superior to his geography achievement. This comparison of Frank's performance in one field with his performance in others is precisely what we want to do in counseling him.

LOCAL NORMS MOST SATISFACTORY

Probably the most satisfactory way to attack the problem of comparability of test results is to establish local norms. Although local norms should be based on at least 100 pupils, confidence in the norms increases with the size of the sample. If we use the same test with several similar groups in successive years, results from these groups may be combined to form the local norm group. For achievement tests and general scholastic aptitude tests, it is usually wise to establish separate norms for each grade level. Graphic methods, involving a minimum of computation, for computing both percentile ranks and standardized scores are described in Appendix B.

Although there is some evidence that the average achievement of girls is slightly superior to that of boys in languages, social studies, and the arts, and the reverse in mathematics and sciences, ordinarily it is not necessary to set up separate norms for boys and girls with this type of test. With clerical or mechanical aptitude tests it is usually more important to establish two sets of norms, one for boys and one for girls, even if we have to combine several grade levels to get large enough groups.

If we cannot set up local norms, we can plan the testing so that the results are comparable. In selecting achievement tests we find that those assembled in batteries have an advantage over most separate tests. Each test in the battery is standardized on the same group of pupils. The scores on each test, then, are all relative to the same base line, namely, the standardization group. Thus, the difference between a pupil's high score in mathematics and his low score in English indicates that there is a real difference between achievement in these two subjects. Our confidence in this conclusion increases if we know that his educational background is similar to that of the group on which the battery was standardized.

Even when separate tests have not been assembled into a battery, some publishers have scaled their separate tests to a single norm group. When this is true we should consider buying several of our tests from the same publisher. It is a wiser purchase, even if we have to spend a little more money, than buying each test from a different publisher. We can examine the difference between a pupil's scores on several tests with more confidence that these differences reflect variations in the pupil's performance and not differences among the norm groups.

NORM SCORES SHOULD BE COMPARABLE

It is helpful also to have the same type of norms available on our tests. A standardized score of 60 is a much higher level of performance than a percentile rank of 60. Table 3 (in Appendix B) shows the equivalent values of percentile ranks and standardized scores in a normal distribution. It may be seen that a standardized score of 60 corresponds to a percentile rank of approximately 84. A percentile rank of 60 is equivalent to a standardized score of only 53. Table 3 enables us to convert the published norms of most tests to a common scale—either percentile ranks or standardized scores. Since there are certain assumptions made in the construction of this table which may not be satisfied by the norm groups for some tests, this practice can be considered a necessary makeshift rather than a recommended procedure.

We have discussed some important characteristics of test information. Now let us look more directly at some of the information which tests can give us.

DETERMINING SCHOLASTIC APTITUDE

One of the main principles of modern education tells us that we cannot expect Leo and Walter to do equally well in school. The achievement of each should be evaluated in terms of his ability. Furthermore, it is not simply that Leo has more ability than Walter, but that he has more of certain abilities necessary for good work in the typical school situation. Application of this principle requires that we evaluate in some way these two boys' ability to succeed in the typical school situation.

Experience has shown that the best single predictor of ability to succeed in future schooling is some measure of past school achievement. This information is, of course, not always available. Originally tests designed to fill this need were called intelligence tests. Since many of such tests have been validated on the basis of their ability to predict success in school, the more descriptive title, scholastic aptitude test, is now common.

Although some of the best scholastic aptitude tests are designed to

be administered to a single individual during an interview, counselors without special training in administering such tests can usually get satisfactory results with paper-and-pencil tests. Group tests are relatively much less expensive of time and money than individual tests because they can be given to about ten primary-grade pupils at one time and to larger groups of older pupils.

INDIVIDUAL TESTS REQUIRE SPECIAL TRAINING

Two scores are ordinarily derived from the results of a scholastic aptitude test, the intelligence quotient (IQ), and the mental age (MA). Leo's mental age is said to be 7 if his raw score on the tests is equal to the average raw score of a norm group of seven-year-olds. Mental age norms are computed by determining the level of difficulty of scholastic material which normal children of any given age can learn.

We may also be interested in how rapidly Leo learns scholastic material. The intelligence quotient attempts to answer this question. If seven-year-old Leo has a MA of 7, he is average, or is said to have an IQ of 100. If he were eight, and had a MA of 7, he would be somewhat below average. His actual IQ is computed by dividing his MA by his chronological age, and multiplying this quotient by 100, in this case 7/8 x 100, or 88. If Leo is 6, his IQ is 7/6 x 100, or 117.

The terms *mental age* and *intelligence quotient* are derived from the older concept of intelligence tests. It is unfortunate that new terms for these measures have not become popular. We shall do well, however, to think of Leo's IQ as a measure of the speed with which he learns typical school material.

Some recent scholastic aptitude tests yield several scores on different types of test material. These tests have been constructed on the theory that better decisions can be made if specific strengths and weaknesses are known. For example, it is more helpful for us to know that Walter has high mathematical ability and low verbal ability than to know simply that Walter's general scholastic ability is average. The research which usually involves such techniques as factor analysis or cluster analysis has been fairly successful in identifying several more or less independent mental abilities. We know that the separate tests of different mental abilities are generally reliable. We know that the correlation coefficients between these tests are fairly low, so they are measuring different aspects of the individual. We are not sure, however, that we know the significance of the various scores. Many counselors are using these newer instruments and depending on their experience with the older general scholastic aptitude tests to help them interpret the results.

SOME SCHOLASTIC APTITUDE TESTS YIELD SEVERAL SCORES

Many paper-and-pencil scholastic aptitude tests require pupils to do considerable reading. This is no criticism of these tests since the typical school situation requires considerable reading, too. The improvement of reading is, nevertheless, one of our schools' accepted objectives. If Leo's reading can be improved, his score on this type of test will be higher.

READING ABILITY MAY AFFECT SCHOLASTIC APTITUDE TEST SCORES

If we have any reason to suspect that Leo's low scholastic aptitude score may be due to poor reading skills, we shall want to retest him with a scholastic aptitude test which requires little reading. Usually these tests are called non-verbal tests. If his score is approximately the same or lower on this second test, there is slight chance that Leo will profit from special instruction to improve his reading. On the other hand, if his score is considerably higher on the non-verbal test, we may be able to increase his scholastic success considerably by directing him into activities aimed at improving his reading skills.

We have recognized the usefulness of scholastic aptitude tests in counseling pupils regarding educational opportunities. Naturally the decisions pupils make regarding the length and content of their formal education will influence the vocational opportunities open to them. The results of scholastic aptitude tests can also be directly useful in counseling pupils regarding their vocational opportunities. Some jobs require a high level of proficiency in reading, writing, speaking, figuring, or other activities typically practiced in school. A pupil's scholastic aptitude is a fair measure of his chances for success in such jobs.

The practice of grouping pupils in classes on the basis of their scholastic aptitude scores is a widely used but questionable one. It is indefensible when it results in the same grouping in all subjects. Unless modifications are made according to subject matter and instructional methods, little is to be gained by this practice, even when different groupings are made for each subject. A better approach is the careful study by each teacher of the child's traits, and a resultant adaptation of content, method, and level of instruction within the class to the needs of each.

CASE-STUDY CLINICS HELPFUL

A counselor is in an excellent position to promote a more rational individualization of instruction, since his job consistently requires him to deal with Harry and Jane rather than with a class. Individuals emerge from a group only as we learn something about them as individuals. Unusual physical characteristics may draw our attention to them first. Later we learn their names. But very few of the complex behavior patterns and personality traits which make up the total unique individual are readily observable within the narrow cultural environment of the average classroom. Still less can we see the basic causes of which many of these traits and much of this behavior are merely symptoms. We shall probably never know the whole child. But the odds are we know things about him that his teachers would profit by knowing. As counselors, then, we can participate actively and constructively in case-study clinics. Such clinics hold more promise for improving the learning opportunities of pupils than have resulted from homogeneous grouping as frequently practiced.

TYPICAL SCHOLASTIC APTITUDE TESTS[2]

AMERICAN COUNCIL ON EDUCATION PSYCHOLOGICAL EXAMINATION FOR HIGH-SCHOOL STUDENTS by L. L. and T. G. Thurstone. American Council on Education, 744 Jackson Place, Washington 6, D. C. This test has four subtests. The first two subtests containing *same-opposite* and *completion* questions are combined to form the L-score. The subtests of arithmetical reasoning and number series form the Q-score. In the 1939 edition of the manual for this test, the authors state: "These two subscores do not represent primary mental abilities, but they represent two groups of abilities significant for curricula that are dominantly linguistic (L-score) or technical (Q-score)." These Quantitative and Linguistic scores when added form the Gross score. This score is comparable to total scores on scholastic aptitude tests. For grades 9-12.

RELIABILITY: Darley states: "For the 1940 edition of the test, hand-scoring edition, Q-score reliability = .94, L-score = .95, Gross-score reliability = .96; for machine-scoring edition, Q-score reliability = .96, L-score reliability = .95, Gross-score reliability = .97. The measures of reliability for the hand-scoring edition were based on scores of 410 freshmen at the Illinois Institute of Technology; reliability for the machine-score edition was computed from the scores of 548 freshmen at the University of Chicago."[3] All coefficients are corrected odd-even. Since 1941, the editions have been constructed so that they are comparable.

VALIDITY: No validity coefficient reported in manual, but believed by Paterson and others to "compare favorably with the best available standard intelligence tests."[4]

---

[2]At the conclusion of the discussion of each type of test, a few tests are described. It is hoped that these descriptions will be valuable to the reader as he reviews them in terms of the discussion. It should not be implied that the tests described are recommended as being better than other tests. The tests were selected for description because they are *typical* of those available. Descriptions of other tests will be found in the books listed in Appendix A.

[3]J. G. Darley, *Testing and Counseling in the High-School Guidance Program* (Chicago: Science Research Associates, 1943), p. 99.

[4]D. G. Paterson, et al., *Student Guidance Techniques* (New York: McGraw-Hill Book Co., 1938), p. 68.

NORMS: A new edition of the test is published each year. Until 1945 percentile norms were issued each spring. At present, schools must determine their norms for editions after 1944, although percentile norms are available for grades 11 and 12 for the 1946 edition.

TIME: 54 minutes required for administration.

COST: Test booklets, per package of 25.....$2.00
Answer sheets, per package of 25..... .50
Specimen set ..... .50
Separate answer sheets must be used for either hand or machine scoring.
Reduction in price for quantity orders of tests and answer sheets.

NEW CALIFORNIA SHORT-FORM TEST OF MENTAL MATURITY (1947 Edition) by E. T. Sullivan, W. W. Clark, and E. W. Tiegs. California Test Bureau, 5916 Hollywood Boulevard, Los Angeles 28, California.

Four subtests are combined to yield the "non-language tests" score and three subtests make up the "language tests." They are combined to yield a "total mental factors" score. The seven subtests are also organized to show abilities in spatial relationships, logical reasoning, numerical reasoning, and verbal concepts. The following forms of the test are available:

Pre-primary.....Kindergarten—Entrance 1st
Primary.....1-3
Elementary.....4-8
Intermediate.....7-10
Advanced.....9-Adult

RELIABILITY: The following split-half reliability coefficients for the mental ages on the various Short-Form Tests are reported in the manual.

Grade	Total	Language	Non-language	N	Grades Tested
Pre-primary	.93	.89	.91	500	1
Primary	.92	.88	.90	700	2-3
Elementary	.95	.95	.91	1,000	4-6
Intermediate	.95	.93	.89	700	7-10
Advanced	.94	.94	.87	400	9-12

Reliability coefficients for the spatial, logical reasoning, numerical, and verbal scores range between .81 and .93 for the above groups.

VALIDITY: Although no figures are reported in the manual, the following statement is indicative: "The traditional method of correlating the results of this series with the averages of several other intelligence tests (protecting results by observing the usual cautions regarding sampling and other statistical safeguards) reveals that the general, or Total Mental Factors I. Q.'s obtained with this test may be used for comparative purposes

with other intelligence tests. However, dealing only with mental ages and intelligence quotients obscures and ignores the separate important factors which constitute mentality; and it is in terms of these factors that the abilities of children should be diagnosed."

NORMS: For non-language, language and total mental factors scores, mental age and grade equivalents are provided in the manual. Percentile norms are provided for all scores at each age. It is possible to compute non-language, language and total I. Q. Percentile rank of I. Q.'s for various populations are given in the manual.

TIME: This test is a power test rather than a speed test, although time limits are provided for the convenience of the examiner. The Short-Form requires one period for administration.

COST: Per 25 tests.....$1.20
Per copy in smaller quantities ..... .10
Specimen set ..... .35

KUHLMAN-ANDERSON INTELLIGENCE TEST by F. Kuhlman and R. G. Anderson. Educational Test Bureau, 720 Washington Avenue, S. E., Minneapolis 14, Minn.

These tests are issued in separate booklets for each grade from 1 through 6, a booklet for grades 7 and 8, and one for grade 9 through maturity.

RELIABILITY: The manual states: "We have attempted to make the tests reliable by adjusting the difficulty of the tests used at each age to the mental development found there. The tests in the scale of 39 tests become progressively more difficult. Each battery of tests presents the same degree of difficulty at the age at which it is used as does any other battery at the age at which it is used.

"We have attempted to make conditions under which children take the tests as uniform as possible by giving preliminary examples for practice for each test, these not being scored, and by giving complete directions for each test, not acquiring or permitting the examiner to supply details according to her own judgment. This tends to eliminate unreliability of test scores that are due to unreliability of the examiner, usually counted in as unreliability of the tests.

"Again, each of the 10 tests in the battery used is scored independently of the rest, and the score earned on the battery is the median of the 10 scores. This eliminates the undue influence of any unusual variation in the score on some particular test at any time." No reliability coefficients are reported in the manual.

VALIDITY: The manual states: "In the present tests, chronological age is used as the criterion of what the tests propose to measure. We propose to measure mental development from the age of 5 to mental maturity. For

this purpose that test is most valid which shows this development best, by having the highest rate of increase in score through successive years. This trait has been called the discriminative capacity of the tests, or the ability to make fine discrimination between small increments in mental development. The age norm table gives a rough indication of the discriminative capacity of each test." No validity coefficients are reported in the manual.

NORMS: The authors recognize that the Mental Age and I.Q. are the commonly used ways of expressing intelligence test results and have provided for this. But they advocate the use of Mental Growth Units and the Percent of Average norms.

TIME: Approximately 45 minutes gross time in grades 9 to maturity; less time in lower grades.

COST: Per package of 25 test booklets for any grade, including key, class record and directions.....$1.35
Specimen set, postpaid..... 1.00

OTIS QUICK-SCORING MENTAL ABILITY TEST by Arthur S. Otis. World Book Company, Yonkers-on-Hudson 5, New York.

Three tests for different grade levels are available. The Alpha Test for the last half of first grade through grade 4; the Beta Test for grades 4 to 9; the Gamma Test for high school through college. Two equivalent forms available for the Alpha Test and four for the Beta and Gamma Tests. The Alpha Test can be administered as a verbal or as a non-verbal test using the same blank.

RELIABILITY:

Alpha Test—Non-verbal .68; Verbal .71. Reliability coefficients obtained by correlating Form A with Form B for tests administered to a single grade.

Beta Test—When Form A was correlated with Form B, the following reliability coefficients were obtained. In all grades Form A was given before Form B. Comparable coefficients obtained when testing order was reversed. Number of pupils used in this study is not indicated in manual. Grade 4 = .73; Grade 5 = .98; Grade 6 = .83; Grade 7 = .71; Grade 8 = .83; Grade 9 = .67. The following corrected split-half reliability coefficients for an unspecified form or number were obtained: Grade 4 = .81; Grade 5 = .92; Grade 6 = .90; Grade 7 = .87; Grade 8 = .86; Grade 9 = .79.

Gamma Test—Split-half coefficients (corrected) for 257 pupils in following grades were: Grade 10 = .90; Grade 11 = .91; Grade 12 = .85.

VALIDITY:

Alpha Test—This test correlated with Primary Examination, another Otis scholastic aptitude, yielded a coefficient of validity. Another indication of validity was obtained by correlation of test scores and grade placement.

These validity coefficients are:

	Verbal	*Non-Verbal*	*Total*
Alpha with Primary Examination	.70	.61	.65
Alpha with Grade Placement	.86	.78	.86

Beta Test—Validity was determined by finding items which differentiated between groups of pupils making rapid progress and those making slow progress through school.

Gamma Test—This was correlated with Higher Examination for grades 10, 11, and 12. The average correlation for these groups, totaling 1,007 pupils, was .86.

NORMS: Grade placement, mental age, and I. Q. norms available for all forms of the test.

TIME: Alpha—20 minutes
Beta—30 minutes
Gamma—30 minutes

COST: Alpha Test:
Per package of 25.....$1.45
Specimen set ..... .35
Beta Test:
Per package of 25, Form A or B 1.10
Per package of 25, Form Cm or Dm 1.20
Specimen set of any form ..... .35
Gamma Test:
Per package of 25, Form Am or Bm 1.20
Per package of 25, Form C or D 1.10
Specimen set of any form ..... .35

THE CHICAGO TESTS OF PRIMARY MENTAL ABILITIES (Single Booklet Edition) by L. L. Thurstone and T. G. Thurstone. Science Research Associates, 228 South Wabash Avenue, Chicago 4, Ill.

This test is designed to measure six important factors or mental abilities. There may be many other abilities, but the Thurstones feel that only these six have been documented for practical use. Each test of these factors yields a separate score. They are *N*, Number; *V*, Verbal-Meaning; *S*, Space; *W*, Word-Fluency; *R*, Reasoning; and *M*, Memory. For ages 11-17.

RELIABILITY: Using groups of approximately 200 pupils, the corrected odd-even reliability coefficients were obtained for each half-year for grades 6, 8, 10, and 12 for the long form of the test. On all factors except M and W, the coefficients were .95 or above for each group. The coefficients for M ranged in the sixties for grades 6 and 8, in the seventies for grade 10, and in the low eighties for grade 12. No reliabilities are available for the Word Fluency section of the test. The "Single Booklet Edition" may not be as reliable.

VALIDITY: The estimated correlations of each of the six composite scores with the primary ability it is intended to appraise are: N .90; W .91; V .97; S .92; M .79; and R .90. The intercorrelations between the factors are low; the median is .39.

NORMS: Percentile ranks and age equivalents for each half-year from 11 to 17½.

TIME: 2 hours. Can be divided into two 1-hour sessions or three 40-minute periods.

COST:

Test booklets (hand scored) per package of 25	$3.75
Profile cards, set of 14	.75
Scoring stencils, set	.75
Memory cards, set of 24	1.00
Extra test manuals, each	.50
Specimen set	2.75

LEE-CLARK READING READINESS TEST by J. M. Lee and W. W. Clark. California Test Bureau, 5916 Hollywood Boulevard, Los Angeles 28, Calif.

Designed to predict readiness to read of children in kindergarten or first grade. Yield three part scores and a total score. Part I, Letter Symbols, is based on tests of matching and crossing out letters. Part II, Concepts, tests vocabulary and ability to follow oral instructions. Part III, Word Symbols, is tested by identification of letters and words which are the same as the printed stimulus words.

RELIABILITY: Corrected split-half coefficients based on 170 entering first-grade pupils.

I.	Letter symbols	.867
II.	Concepts	.832
III.	Word symbols	.936
	Total for test	.925

VALIDITY: Obtained by finding the correlation of this test with a test designed to measure reading achievement. For one group of 72 first-grade pupils, the correlation between Lee-Clark Reading Readiness Test, given at the beginning of the year, and Lee-Clark Reading: Primer, given after nine months of instruction, was found to be .67. In another group of 374 above-average-in-ability pupils, the correlation was .43.

The correlation between the Lee-Clark Reading Readiness Test and the California Test of Mental Maturity, Pre-Primary Series for a group of 377 first-grade pupils was found to be .65.

NORMS: Grade-placement equivalent, descriptive classification from very low to high, and probable percent of failure for each score are included in the manual.

TIME: Approximately 30 minutes, preferably divided into two sessions on the same day.

COST:

Per 25 tests, package	$1.20
In smaller quantity, each	.10
Specimen set, each	.35

METROPOLITAN READINESS TEST by G. H. Hildreth and N. L. Griffiths. World Book Company, Yonkers-on-Hudson 5, New York.

A test to determine the readiness of children to do first-grade work in reading and numbers. Correlates highly with general intelligence tests.

RELIABILITY: Not reported in manual.

VALIDITY: No validity coefficients reported in manual. Data from one school for 494 pupils show considerable correspondence between scores on this test given at the beginning of the first grade with scores on achievement tests given at the end of the first grade.

NORMS: Percentile ranks based on 10,449 entering first grade and percentile norms for ages 5½ through 7 are available.

TIME: 70 minutes. Authors recommend that it be divided into several testing periods to lessen fatigue.

COST:

Per package of 25	$1.50
Specimen set	.35

MEASURING ACHIEVEMENT

The cumulative record usually contains some items indicative of the pupil's scholastic achievement. Teachers' marks are one indication. Before they are accepted at face value, it is well to discover the marking policy of the school. Unless there is evidence that a clearly stated marking policy is conscientiously followed by teachers, marks are usually not very good measures of pupil achievement.

Another indication is found in achievement test results. There are many varieties of achievement. Since the acquisition of information is one of the objectives of nearly all school subjects, it is only natural that most achievement tests attempt to measure how much of this information each pupil has learned. Tests which are largely informational in character must be carefully checked. They are valid measures of achievement only to the extent that the contents of the tests are an adequate sampling of all the information pupils have had an opportunity to learn.

Some achievement tests are designed to identify pupils who have not mastered the skills basic to further progress in school. For example, knowing the meaning of numbers is a skill basic to successful achievement in arithmetic. Sometimes these tests are called diagnostic tests. Frequently some parts of an achievement test battery are devoted to this type of testing.

In recent years a number of test-makers have been concerned with developing tests which cut across traditional subject-matter lines. These tests of general educational development or proficiency cover large areas, such as social studies, mathematics, and natural sciences. Such tests usually put little emphasis on testing the range of pupils' information. Rather, the tests attempt to measure the pupils' ability to apply what information they have in the solution of new problems. Or the pupils are asked to interpret or evaluate unfamiliar material. While these newer tests hold some promise for evaluating the more permanent results of schooling, little is known of the usefulness of such tests for guidance purposes. They appear to have considerable validity for predicting future achievement in broad scholastic fields, but no conclusive statistical evidence on this point is available.

NEWER ACHIEVEMENT TESTS LESS FACTUAL

TYPICAL ACHIEVEMENT TESTS

COOPERATIVE GENERAL ACHIEVEMENT TESTS by M. Willis, E. Spaney, R. E. Watson, and others. Cooperative Test Service of the American Council on Education, 15 Amsterdam Avenue, New York 23, N. Y.

These tests are issued in separate booklets for each of the following fields: Part I, social studies; Part II, natural science; and Part III, mathematics. Forms N, O, and P are intended as survey measures of the various high-school courses in each field and are divided into several subject-matter divisions. The items were selected to cover those aspects of each subject which might be considered of lasting significance. The Revised Series, Forms OR, S, and T, instead of items dealing with the topical content of the field, are divided into two parts: The first calls for a knowledge of the terms and concepts essential to an understanding of the field; the second tests the pupil's ability to comprehend and interpret typical materials in the field. Designed for use with grades 10 through 12 and freshmen entering college.

RELIABILITY: Not reported in manual.

VALIDITY: Items were selected by experts and difficulty of words checked against Thorndike's Word Book. Validity for these tests can best be determined in the light of the objectives of the local school.

NORMS: Percentile norms, based on the total Scaled Score for each test are provided for end-of-year high-school students at each grade level and for entering college freshmen.

TIME: 40 minutes for each test. 2 hours for battery of 3 tests.

COST: Price for each test of the General Achievement Tests. Each of the 3 parts (Test I, II, III) must be ordered if the complete battery is desired.

Test books, per copy	$0.07
Answer sheets (for use when machine-scoring or re-use of booklets is planned)	.015
Reduction in price for quantity orders of tests and answer sheets.	
Specimen set containing one copy each of Parts I, II, and III with necessary materials	.50

METROPOLITAN ACHIEVEMENT TESTS (Rev. Ed.) by R. D. Allen, H. H. Bixler, W. L. Connor, F. B. Graham, and G. H. Hildreth. World Book Company, Yonkers-on-Hudson 5, New York.

Designed to measure achievement in subjects taught in grades 1 through 8. The Primary I Battery contains tests of word and phrase recognition, word meaning, and numbers suitable for grade 1. Tests of Reading, Vocabulary, Arithmetic Fundamentals and Problems, and Spelling comprise the Primary II Battery. The Elementary Battery has the same tests and an additional test of Language Usage. The Intermediate (grades 5-6) and Advanced (7, 8, and beginning of grade 9) partial batteries include tests of reading, vocabulary, arithmetic fundamentals and problems, spelling, and English. The complete batteries contain, in addition, tests of literature, history and civics, and geography. Two forms of tests are available.

RELIABILITY: Corrected split-half reliabilities range from .800 to .970.

VALIDITY: Based upon examination of textbooks and courses of study. Validity of test as measure of achievement of current instruction must be determined locally.

NORMS: Scores may be expressed as grade or age equivalents or percentile ranks.

TIME:

Primary I Battery	1 hour
Primary II Battery	1 hour 25 minutes
Elementary Battery	2 hours 15 minutes
Intermediate—Complete Battery	3 hours 20 minutes
Advanced—Complete	3 hours 40 minutes
Intermediate—Partial	2 hours 40 minutes
Advanced—Partial	2 hours 40 minutes

COST:

Primary I Battery:	(Per package of 25)	$1.60
	(Specimen set)	.35
Primary II Battery:	(Per package of 25)	1.65
	(Specimen set)	.35
Elementary Battery:	(Per package of 25)	2.25
	(Specimen set)	.35
Intermediate or Advanced Complete Battery:	(Per package of 25)	2.70
	(Specimen set)	.35

Partial Battery: (Per package of 25) 2.20
(Specimen set) ..... .35

Certain tests are published separately. See publisher's catalog for list and cost.

PROGRESSIVE ACHIEVEMENT TEST by Ernest W. Tiegs and Willis W. Clark. California Test Bureau, 5916 Hollywood Boulevard, Los Angeles 28, Calif.

Designed to measure and analyze the status of pupils in reading, arithmetic, and language skills. The test is organized to provide scores for reading vocabulary, reading comprehension, arithmetic reasoning, arithmetic fundamentals, and language. The tests are available as a battery and as separate booklets for reading, arithmetic, and language in the following forms:

	Grades
Primary Battery, Forms A, B, and C	1-3
Elementary Battery, Forms A, B, and C	4-6
Intermediate Battery, Forms A, B, and C	7-9
Advanced Battery, Forms A and B	9-14

RELIABILITY: Coefficients of reliability, obtained by giving alternate forms, are reported by the publisher for typical grades, as follows:

Subject	Primary Grade 3	Elementary Grade 5	Intermediate Grade 8	Advanced Grade 10
1	2	3	4	5
Reading vocabulary	.89	.88	.90	.90
Reading comprehension	.92	.93	.89	.89
Total reading	.93	.93	.92	.92
Arithmetic reasoning	.84	.89	.92	.88
Arithmetic	.86	.96	.95	.92
Total arithmetic	.88	.95	.95	.93
Language	.93	.91	.94	.93
Total for test	.96	.97	.97	.98

VALIDITY: According to the manual, "The content is based on some of the most tangible and most easily identified objectives of the curriculum . . . The selection of items was based on careful study of the curriculum objectives of the progressive city and State courses of study . . . The tests were tried out in widely separated geographical areas . . . Studies have been made of individual items under a variety of conditions."

NORMS: According to the publisher's catalog, the standardization of the test has been based on more than 50,000 cases at each level. Both age-grade and percentile norms are provided in the manual.

TIME: The time required for the complete battery of five tests is approximately as follows:

Primary ..... 1 hour 30 minutes
Elementary ..... 2 hours

Intermediate ..... 2 hours 30 minutes
Advanced ..... 2 hours 30 minutes

COST:

	Primary A, B, or C	*Other Batteries A, B, or C*
Battery, per 25 tests	$1.75	$1.90
Per copy	.10	.10
Reading, per 25 tests	1.20	1.20
Per copy	.10	.10
Arithmetic, per 25 tests	1.20	1.20
Per copy	.10	.10
Language, per 25 tests	.90	.90
Per copy	.10	.10

Specimen set, any item—35 cents

STANFORD ACHIEVEMENT TEST by T. L. Kelley, G. M. Ruch, and L. M. Terman. World Book Company, Yonkers-on-Hudson 5, New York. There are three different complete batteries for grades 2-9, and two different partial batteries from grades 4-9. The Primary Battery for end of grade 2 and 3 contains tests of paragraph meaning, word meaning, spelling, arithmetic reasoning, and arithmetic computation. The partial Intermediate Battery for grades 4-6, and the partial Advanced Battery for grades 7-9 contain a test of language usage in addition to those in the Primary Battery. The complete Intermediate and Advanced Batteries have the six tests of the partial batteries and additional tests in literature, social studies I (history primarily), social studies II (geography primarily), and elementary science. Five comparable forms of each battery available.

RELIABILITY: For a group of 226 pupils in grade 5, the corrected reliability coefficient ranged from .71 for Social Studies I to .94 for Spelling. The reliability for the complete battery with this group was found to be .97. In a sample containing 146 8th-grade pupils, the reliabilities of subtests was found to range from .74 to .93, with reliability of total battery of .97. The Primary Battery reliabilities range from .86 to .95, with total reliability of .97 for 164 pupils in grade 3.

VALIDITY: Items based on analysis of representative courses of study, evaluation by subject-matter specialists, and try-outs in widely separated schools. Validity of test as measure of achievement of current instruction must be determined locally.

NORMS: Two types of grade- and age-equivalent norms are available: (1) norms based on groups from which accelerated or retarded pupils are removed, and (2) traditional norms based on the total population tested.

TIME:

	Approximate Working Time
Primary Battery	1 hour 5 minutes
Intermediate Battery—Complete	2 hours 30 minutes
Advanced Battery—Complete	2 hours 30 minutes
Intermediate Battery—Partial	1 hour 50 minutes
Advanced Battery—Partial	1 hour 50 minutes

COST:

Primary Battery	Intermediate or Advanced—Complete
Per package of 25...$1.35	Per package of 25...$2.70
Specimen set ..... .35	Specimen set ..... .35

Intermediate or Advanced—Partial
Per package of 25 .....$2.20
Specimen set ..... .35

The Intermediate and Advanced Batteries are available in machine-scoring edition with separate answer sheets. Certain of the subtests can be purchased separately. For further information and prices, consult the publisher's catalog.

IOWA TESTS OF EDUCATIONAL DEVELOPMENT by K. W. Vaughn, J. Peterson, T. W. Naucker, and P. Blommers under the direction of E. F. Lindquist. Science Research Associates, 228 South Wabash Avenue, Chicago 4, Ill.

This test is designed to measure the general educational background and development of individual pupils. Two equivalent forms are available. The test is for grades 9-13, inclusive. The nine subtests of the battery are entitled: Understanding of Basic Social Concepts, Background in the Natural Sciences, Correctness in Writing, Ability to Do Quantitative Thinking, Ability to Interpret Reading Materials in the Social Studies, Ability to Interpret Reading Materials in the Natural Sciences, Ability to Interpret Literary Materials, General Vocabulary, and Use of Sources of Information.

RELIABILITY: The authors report that "the reliability coefficient for each of the Iowa Tests of Educational Development is close to .91."

VALIDITY: The validity of this type of test can best be determined in terms of the local situation.

NORMS: Percentile ranks by half-years for grades 9 through 12.

TIME: A minimum of 3 half-days, 7 hours 11 minutes of which is actual working time.

COST: Testing materials, scoring service, and individual and school summary profiles furnished for 75 cents per pupil. Transportation costs extra.

UNITED STATES ARMED FORCES INSTITUTE TESTS OF GENERAL EDUCATIONAL DEVELOPMENT. Distributed by Cooperative Test Service of the American Council on Education, 15 Amsterdam Avenue, New York 23, N. Y., and Science Research Associates, Inc., 228 S. Wabash Avenue, Chicago 4, Ill.

These tests are issued in five separate booklets:

Test 1—Correctness and Effectiveness of Expression
Test 2—Interpretation of Reading Materials in the Social Studies
Test 3—Interpretation of Reading Materials in the Natural Sciences
Test 4—Interpretation of Literary Materials
Test 5—General Mathematical Ability

Tests 2, 3, and 4 measure the pupil's ability to comprehend, interpret, and critically evaluate typical materials in each area. Test 5 involves largely arithmetical problem-solving. Test 1 covers spelling, punctuation, capitalization, usage, and sentence structure. Form B available.

RELIABILITY: Not reported.

VALIDITY: Many colleges are indicating their belief in the predictive power of these tests by admitting students on the basis of their scores. Since the tests are not designed as end-of-course achievement examinations, validity can best be determined in terms of local objectives.

NORMS: Standard scores and percentile norms on each test are available, based on 35,432 high-school seniors tested just prior to being graduated from a general high-school curriculum.

TIME: Non-timed; approximately 2 hours for each test.

COST: Test booklets, per package of 25 .....$2.00
Answer sheets (separate answer sheets must be used)
Either hand-scoring or machine-scoring answer sheets per package of 25 ..... .65
Reduction in price for quantity orders of tests and answer sheets, Specimen set (each test) ..... .50

INTEREST TESTS

How Margaret feels about her job is an important factor in how well she will do on it. Her ability to do the job well may count for very little if she has no interest in it. Successful placement in or out of school implies that Margaret is engaging in some activity in which she is both interested and capable. Margaret's interests change. Activities which she liked in the sixth grade no longer interest her now that she is a high-school senior. Her interests will continue to change, more slowly perhaps, depending on her adult experiences. Common experience, however, leads psychologists to believe that certain aspects of interests are remarkably persistent and stable. It is in the search for these more permanent aspects of interests, particularly those important for adjustment in various vocational fields, that interest tests have been developed.

The odds are high that Margaret may have a quite different vocational goal in the twelfth grade from that which she had in the sixth. In fact, she may reach her last year in high school with no vocational aim at all, or even with an aim entirely out of line with her ability, training, or opportunities. Margaret may not recognize her interests. Her friends, relatives, or teachers may confuse her by snap judgments of her interests based on limited observation of her activities. We cannot take seriously the vocational choice indicated in Margaret's cumulative record. Anecdotal records may be more helpful, as may be the list of extracurricular activities in which she has engaged. We know little about the validity of these data in predicting future vocational adjustment. We do know that such data are not highly reliable.

Reliability coefficients of .85 and higher are reported in the manuals of several interest tests. Whatever these tests are measuring, they are measuring with a fair degree of reliability. Correlations between ability and interest, on the other hand, are surprisingly low. Margaret is sure to show interest in some activities in which she can engage with only a fair degree of success, and little interest in other activities in which she has considerable ability. Measures of interest or motivation are not good measures of information and ability.

INTEREST AND ABILITY NOT CLOSELY RELATED

The types of information we can get from interest tests vary a great deal. Dr. Strong has scored his men's test for thirty-nine specific occupations and his women's for twenty-five. Most of these occupations are on a professional level. Strong himself does not recommend the test for boys and girls under seventeen. Many of the interest inventories designed for use in high schools gives scores in broad fields of interest, such as mechanical or computational rather than scores in specific occupations. The correlation between scores in these broad areas of interest is usually low, although some test-makers give little or no information on this important point. Study of the pattern of Margaret's interests is more helpful than simply noting the area of her highest interest. A strong scientific interest coupled with a secondary interest in mechanical activities would be interpreted one way, while if her mechanical interest is low and her computational score high, quite a different interpretaton would be made.

In short, interest tests can give important information about Margaret that would not be otherwise available. The more mature Margaret is, the better chance we have of discovering those interests which will be important in her job adjustment. But we must not assume that interest in an occupational field indicates either the ability or opportunity to enter that field. The most unfortunate trend in the whole area of guidance testing has been the tendency of some counselors to overemphasize interest test results. To counsel pupils on the basis of an interest test with little regard for other pertinent items in the individual inventory is worse than useless. It fosters disillusionment and frustration in those whose abilities are not in line with the interests we have encouraged. Interests must be recognized, but they are not the whole show.

INTEREST TESTS NEED SUPPORTING DATA

TYPICAL INTEREST TESTS

BRAINARD OCCUPATIONAL PREFERENCE INVENTORY by P. P. and R. T. Brainard. Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

A revision of Specific Interest Inventory containing 140 items. The testee indicates "Like-Dislike" on a 5-point scale to each item. Yields scores in 28 occupational sections. These sectional scores are combined to indicate interest in the following seven fields: Commercial, personal service, agriculture, mechanical, professional, esthetic, and scientific. The testee can score and prepare the profile. Suitable for high school and above.

RELIABILITY: The authors report a reliability of .81 computed by Ghiselli's method. They believe "that the true reliability is higher than this."

VALIDITY: Inventory is constructed in a large measure on the basis of experience with the Specific Interest Inventory. No coefficients reported in manual.

NORMS: Separate norms for adult men, adult women, high-school boys, and high-school girls.

TIME: Untimed, but 30 minutes is usually sufficient.

COST: Inventory booklets (reusable) .....each $ .25
Record form (answer sheet) per package of 25 ..... 1.25
Specimen set ..... .50
Reduction in price for quantity orders. See publisher's catalog.

KUDER PREFERENCE RECORD by G. F. Kuder. Science Research Associates, 228 South Wabash Avenue, Chicago 4, Ill.

This test yields scores for nine areas of interests. They are mechanical, computational, scientific, persuasive, artistic, literary, musical, social service, and clerical. The pupil indicates which of three activites he likes most and which he likes least. The test was constructed so that the scores in each of the nine areas are relatively independent of each other. Thus, when plotted on the profile furnished with the test, they form a basis for evaluating the relative strength of interests in these areas.

RELIABILITY: From a number of studies, the average reliability for all scales is about .90.

VALIDITY: When interest scores of men and women engaged in various

occupations are compared with a base group, the significant differences found are frequently in agreement with logical expectations.

NORMS: Separate norms for male and for female high-school students and adults. College norms are being developed. An equation for computing masculinity-femininity score presented in the manual. A specific occupational score can be computed for accountant-auditor. No other specific occupational score equations are available at present. Mean profiles for some occupations are provided.

TIME: Untimed. Usually requires 30 to 40 minutes.

COST:

Hand-scored (Form BB)	
Booklet (reusable) with one answer sheet	$ .48
Extra answer pads, per package of 25	2.00
Specimen set	.75
Machine-scored (Form BM)	
Booklet (reusable)	.35
Answer sheet, per package of 100	2.35
Machine-scoring keys, per set	7.50
Profile sheets, per package of 25	.50

OCCUPATIONAL INTEREST INVENTORY by Edwin A. Lee and Louis P. Thorpe. California Test Bureau, 5916 Hollywood Boulevard, Los Angeles, Calif.

Scores are obtained for six areas; namely, Personal-Social, Natural, Mechanical, Business, the Arts, and the Sciences. Three additional scores indicate types of interests; they are verbal activities, manipulative activities, and computational activities. A final score reveals the level of interests. The Intermediate Form is designed for pupils in junior high school or above. The Advanced Form of this test is for senior high, college, and adult levels. A profile for recording scores is on the front of the test booklet. The specific items of the Advanced Form are coded according to the *Dictionary of Occupational Titles*.

RELIABILITY: Test-retest reliabilities are reported to be .88 to .93. Testing administered over a period of 4 weeks.

VALIDITY: Authors report no validity coefficients. They state that the following factors were considered in the construction of the inventory to make it more valid: Selection, design, balance, and presentation of items. A comprehensive guidebook entitled *Occupational Selection Aid* (published as a supplement to the Advanced Form) provides for a classification of over 500 specific job titles listed according to 21 interest pattern groups.

NORMS: Percentile norms for both forms of the test are provided separately for males, for females, and for males and females combined.

TIME: Untimed. Usually requires 30 - 40 minutes.

COST:

Either form, per package of 25	$1.75
Specimen set	.25

VOCATIONAL INTEREST BLANKS (for Men or Women) by E. K. Strong, Jr. Stanford University Press, Stanford, Calif.

Designed to reveal the extent to which a testee's interests agree with those of persons engaged in certain occupations. The men's blank may be scored for 39 specific occupations, 6 occupational groups, and 3 special variables; namely, Interest Maturity, Occupational Level, and Masculinity-Femininity. The women's blank may be scored for 25 specific occupations. Probably not suitable for pupils under 17 years.

RELIABILITY: For all scales the average reliability is about .88.

VALIDITY: The various scoring keys were developed by testing persons engaged in various occupations.

NORMS: Scores for each of the scales are based on the standardization group.

TIME: Untimed.

COST:

Blanks per package of 25	$2.00
Separate scoring key for each occupation, each	1.00
Answer sheets (use optional), per package of 25	.75
Specimen set	.15
Reduction in price for quantity orders.	

SCORING: Scoring by hand is laborious. Machine scoring is available at various psychological centers throughout the country. Usual cost is about $1.50 for each men's blank and $1.00 for each women's blank if all scales are scored. Less expensive if representative scales are selected. A recently invented machine is used by Engineers Northwest, 314 Second Avenue South, Minneapolis, Minn.

VOCATIONAL INTEREST INVENTORY by C. E. and E. G. Germane. Contained in the book *Personnel Work in High Schools* by the same authors. Silver Burdett Company, 45 East 17th Street, New York City, N. Y.

This test yields scores in nine areas; namely, commercial, mechanical, esthetic, manual, agricultural, academic (professional), scientific (professional), general service, and domestic. The subject rates his liking or distaste for 35 activities in each of the nine areas.

RELIABILITY: No published studies. Current studies under way by certain Supervisors of Occupational Information and Guidance indicate satisfactory reliability.

VALIDITY: No published studies.

NORMS: Based on studies made by the Germanes in 50 Missouri high schools.

TIME: Untimed.

COST: Silver Burdett state: "We have granted the purchasers of this book the right to reproduce and utilize these tests in his or her own school system without charge. We do not provide and offer for sale printed copies of these tests."

JUDGING PERSONAL ADJUSTMENT[6]

To what extent can tests be used in judging personal development? Most cumulative records contain some information about personality or character, as distinguished from scholastic abilities or achievement and interests. These data are frequently in the form of a rating by one or more teachers of each pupil on a series of personality traits, such as industry, dependability, sociability, leadership, and cooperation. The list of these so-called traits could be expanded indefinitely. What do these ratings tell us about Philip that we cannot already discover from other data in the record? In the first place, unless unusual care is taken in the construction and marking of the rating scale, repeated investigations have shown that the results will be very unreliable. Can we say Philip is generally industrious or generally lazy? Is he not industrious at some jobs and lazy at others? Is he not a leader on the athletic field and a follower in his class meetings?

The difficulty of defining a few general personality traits which Philip exhibits under most circumstances is a problem which plagues the personality test-maker as well as the rater. The list of traits measured by personality tests grows as long as the list of traits measured by rating scales. Some of the problems involved in personality evaluation depend on the type of test used.

One method being used by an increasing number of psychologists involves exposing Philip to some more or less vague stimulus to which he is given considerable freedom in responding. The projective techniques may vary from the free-word-association test to Philip's account of everything he sees in a series of ink blots, or the stories he tells when presented with a series of pictures. Since Philip's responses are largely uncontrolled, these techniques require considerable time and extensive training to score and interpret. Few schools have personnel with either the training or the time to administer projective tests. The interpretation of the scores obtained is even more difficult. It appears likely that unless major modifications are made in these techniques, they will remain the tools of the skilled clinical psychologist. Certainly few counselors can justify including such tests in their general program.

PROJECTIVE TECHNIQUES ARE TOOLS OF SKILLED PSYCHOLOGISTS

[6] For a review of research in this area consult an article by Elbert Ellis, "The Validity of Personality Questionnaire," *Psychological Bulletin*, September, 1946.

The most common method of evaluating personality is by means of paper-and-pencil tests. These tests require Philip to answer questions about how he feels or acts in certain situations. Does he think most people regard him as queer? Does he love his mother more than his father? Does he think some of his teachers are too sarcastic? Usually several scores are obtained from such tests. These scores may attempt to evaluate Philip's success in adjusting to his home, the school, or his classmates. Or they may indicate tendencies toward emotional instability, lack of self-confidence, excessive day-dreaming, and the like.

An obvious objection to this type of test is that Philip may not be willing to answer such questions truthfully. If he answers the way he thinks he should answer them instead of the way he really feels, we may get quite a false picture of Philip. There is no easy solution to this problem. If we can gain Philip's confidence so that he will respond honestly to the questions, we can get valuable information from these tests. Paper-and-pencil tests of personality may be used with individuals with whom we have established rapport in the individual interview. We do not ordinarily have occasion to use them with large groups.

RAPPORT NECESSARY FOR PAPER-AND-PENCIL TESTS

The best single device available for gathering data on this aspect of pupil behavior is the anecdotal record. All members of the faculty should be encouraged to study and to use this technique. How apparent discrepancies in test or other data may be used to identify pupils with personality or adjustment problems is discussed at more length in Chapter V.

TYPICAL PERSONAL ADJUSTMENT TESTS

THE ADJUSTMENT INVENTORY (Student Form) by Hugh M. Bell. Stanford University Press, Stanford, Calif.

This inventory indicates the testee's home, health, social, and emotional adjustment. Scores in these areas are added to obtain a total score. Contains 140 questions which can be answered *Yes*, *?*, or *No*.

RELIABILITY: Corrected odd-even reliabilities range from .80 to .89 for part scores and .93 for the total score.

VALIDITY: Correlations with other personality tests range from .72 to .90. Statistically significant differences in scores obtained were found between well-adjusted and poorly adjusted groups.

NORMS: Tentative norms published in 1934 are available for high-school men (161), high-school women (190), college men (171), and college women (243). Numbers in parentheses are the number of cases used to establish norms.

TIME: Untimed. Ordinarily 25 minutes is sufficient.
COST: Inventory, per package of 25 ..... $1.75
Specimen set, each ..... .15
Reduction in price for quantity orders.

WASHBURNE SOCIAL-ADJUSTMENT INVENTORY (Thaspic Edition) by J. N. Washburne. World Book Company, Yonkers-on-Hudson 5, New York.

A series of 122 questions, most of which are answered by *yes* or *no* comprise this inventory. It is designed to yield the following scores: Truthfulness, Happiness, Alienation, Sympathy, Purpose, Impulse-Judgment, Control, and Wishes. Only one form of the test. Suitable for junior high school and above. Available for hand or machine scoring.

RELIABILITY: Part scores range from .73 to .88. Total adjustment score .92. Reliabilities were computed by retesting students one semester after they took the first test.

VALIDITY: The manual reports that a bi-serial coefficient of validity of .90 was found by testing 400 pairs matched in age, intelligence, and sex, but contrasted in adjustment.

NORMS: Percentile norms for each sub-test and the total score are given for junior high school, high-school, and for college students.

TIME: Untimed. Usually 30 - 50 minutes.
COST: Inventory, per package of 25 ..... $1.60
Manual for interpreting, each ..... .20
Specimen set (does not include above manual), each ..... .35

SPECIAL APTITUDE TESTS

Finally, tests have been developed to assist us in predicting pupils' probable success in specific school training and in certain vocational fields. Aptitudes for clerical, mechanical, musical, and artistic training or work have been shown not to be highly related to general scholastic aptitude. It has sometimes been falsely assumed that low scholastic aptitude implied high aptitude in one or more of these areas. Human abilities just do not operate that way. If Eddie's general scholastic ability is very low, we cannot predict from that fact what his special aptitudes will be. There is a fair chance that he has not exceptionally high aptitude for any of these vocational fields. But Eddie may be the exception even to this rule. General scholastic aptitude tests give us little or no help in counseling pupils regarding their aptitude for activities which do not involve the three R's.

Aptitude tests for specific high-school subjects have been largely confined to mathematics, foreign languages, and commercial subjects. The results indicate that we can do about as good a job in counseling by using

a general scholastic aptitude test, plus a measure of past achievement in closely related subjects, as by using aptitude tests. It seems inadvisable to use tests of artistic or musical apitude with groups.

APTITUDE TESTS FOR SCHOOL SUBJECTS

In selecting such tests for individual use, we need to search for some evidence that the test measures reliably what we want to know. Few tests in these areas have been published within the last ten years. A few professional organizations, such as those of the engineers and physicians, have been active in developing tests in their fields. These instruments are designed to measure the student's ability to succeed in the training for these professions. They are generally not designed for or available to high-school undergraduates.

Previously we noted that the titles of tests do not always give an adequate description of what a test measures. This caution is particularly applicable to special aptitude tests. Some clerical aptitude tests are excellent predictors of ability to file correctly, but relatively useless in predicting ability to succeed in other activities frequently associated with clerical work. Tests of mechanical aptitude are even more varied. There are tests of mechanical information in which the pupil may be asked to identify a wide variety of tools, or recognize parts of common mechanical contrivances. Some mechanical aptitude tests require pupils to look at drawings of geometric figures and decide how they fit together to form a certain pattern, or otherwise demonstrate a high degree of spatial perception. There are tests of mechanical comprehension in which the pupil must predict how one part of a mechanical device causes the specified action of another part. Some mechanical aptitude tests measure how fast a pupil can put three dots in a circle, assemble a doorbell, rotate pegs in a pegboard, or do some other task requiring dexterity. Since none of these tests correlates very highly with the others, we cannot assume that any one of them tests all aspects of a pupil's mechanical aptitude.

TESTS OF CLERICAL AND MECHANICAL APTITUDE

In many schools the opportunities for clerical and mechanical training are relatively extensive. Certainly the post-school vocational opportunities in these two fields are large. We should seriously consider including clerical and mechanical aptitude tests in our local testing program. They may well fill an important gap in the data used for counseling.

TYPICAL TESTS OF CLERICAL APTITUDE

MINNESOTA VOCATIONAL TEST FOR CLERICAL WORKERS arranged by D. M. Andrew under the direction of D. G. Paterson and H. W. Longstaff. Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

This test is designed to measure speed and accuracy in checking names and numbers. The first part consists of 200 pairs of numbers. If they are the same, a check mark is placed on a line connecting both numbers. If they are not identical, no mark is made. The second part consists of 200 pairs of names which are checked in a similar manner. For use with high-school pupils or adults.

RELIABILITY: Corrected odd-even reliability is about .90. Test-retest reliability coefficients range from .85 to .91.

VALIDITY: Validation data based on employed clerical workers and high-school commercial students are reported in manual. Coefficients range from about .30 to .65, depending on group and criterion.

NORMS: Percentile norms based on groups of employed men and women clerical workers are given in manual. The manual includes norms for grades 8 - 12.

TIME: 15 minutes.

COST: Per package of 25 ..... $1.25
Reduction for quantity orders
Specimen set ..... .25

TEST OF CLERICAL COMPETENCE by A. J. Cardall and J. G. Hench. Science Research Associates, 228 South Wabash Avenue, Chicago 4, Ill. Designed to measure aptitude for clerical or other occupations in which perceptual ability and ability to deal with small details are important. It consists of four parts: Checking numbers, checking names, verbal classification, and numerical classification. For use in grades 11 and 12, and with adults.

RELIABILITY: Authors report coefficients of .90 to .98 for part scores and .99 for total score using Kuder-Richardson formula for reliability.

VALIDITY: Type of items included based on author's job analysis of clerical occupations.

NORMS: Norms for employed workers and for high-school pupils are available.

TIME: 23 minutes.

COST: Package of 25 ..... $2.55
Specimen set ..... .50

TYPICAL TESTS OF MECHANICAL APTITUDE

REVISED MINNESOTA PAPER FORM BOARD TEST by R. Likert and W. Quasha. Psychological Corporation, 522 Fifth Avenue, New York 18, New York.

This test is designed to measure the ability to visualize and manipulate mentally geometric forms. The authors state: "High scores on this test

are predictive of (1) ability to learn mechanical drawing and descriptive geometry; (2) success in mechanical occupations; and (3) success in engineering courses." Suitable for age 9 or older. Two forms available in either hand-scored or machine-scored editions.

RELIABILITY: If only one form is given, the manual reports a reliability of .85. If both forms are given, the reliability is .92.

VALIDITY: This is the only paper-and-pencil test in the Minnesota Mechanical Ability Battery which correlated satisfactorily with a quality criterion of mechanical ability. This validity coefficient was found to be .52 and when corrected for attenuation .61. Other evidences of validity are reported in the manual.

NORM: Percentile rank:
Males and females separately for:
Ages 9, 10, 11, 12, 15, 16-25, 25-60, 4th grade, 5th grade.
Males only for:
High-school seniors, high-school graduates, liberal arts college freshmen, and engineering school students by year of study, printers' apprentices, first-year vocational school students, and junior and senior vocational school students.

TIME: 20 minutes.

COST: Hand-scored edition, per package of 25 ..... $1.25
Machine-scored edition, per package of 25 ..... 1.55
Machine-scored answer sheets, per package of 50 ..... 1.50
Specimen set ..... .50

TESTS OF MECHANICAL COMPREHENSION by G. K. Bennett and D. E. Fry. Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

Designed to measure the capacity of an individual to understand various types of physical relationships. The ability measured is believed to be important in physics courses, in many trade school courses, and in engineering schools. Form AA is suitable for high-school and adult men with comparable education. Form BB, more difficult than Form AA, is suitable for male candidates to engineering schools, engineering students, and adult men of comparable education. Form W1 is the women's form of the series; the difficulty is between that of AA and BB.

RELIABILITY: The Form AA split-half reliability coefficient corrected for a group of 9th-grade boys was found to be .84. Form BB was found to have a corrected split-half reliability of .80 with a group of college freshmen engineers. The corrected split-half reliability of Form W1 for a group of enlisted WAVES was found to be .77.

VALIDITY: Form AA was found to correlate .5 with average grade in

military technical courses. Other data reported in manual showing similar relationships between scores and various occupational and education criteria.

NORMS: Form AA has norms based upon high-school students by grades, engineering school freshmen, candidates for defense training courses, and other groups. Form BB norms are based on engineering school applicants, engineering school freshmen, and other groups. Form W1 has norms for freshmen and senior high-school girls, candidates for mechanical courses, employees at light mechanical work, and WAVES enlisted personnel.

TIME: No time limit. Usually takes about 30 minutes.

COST: Test booklets (reusable)
Single copy ..... $ .15
Per package of 25 ..... 3.00
Answer sheets (even if the test is hand scored, answer sheets must be used.)
Per package of 50 ..... 1.50
Specimen set for any one form ..... .30
Reduction in price of booklets and answer sheets on quantity orders.

Chapter IV

ADMINISTERING, SCORING, AND RECORDING RESULTS OF TESTS

TEST RESULTS ARE worse than meaningless if there have been errors in administering and scoring the tests, or if there has been inaccurate recording of results. Scoring and recording of most standardized tests are routine clerical jobs, although a few tests require judgment on the part of scorers. The administration of group tests, however, is a professional activity. Usually good teachers can, through training and experience, become good examiners. There are exceptions to this rule, since some excellent teachers are temperamentally incapable of restricting their test-room activities within the bounds set by the test manual. On the other hand, poor teachers are almost invariably poor examiners.

In schools which do not have separate testing divisions, teachers usually are the logical persons to administer group tests. Other members of the staff may assist. Not all teachers should be included, but teachers can constitute the main body of the school's examiners. Ordinarily, the counselor, by reason of special training, will be in the best position to administer tests to individuals. Sometimes, however, we must consider the possibility that some other member of the staff may be able to do a better job than the counselor of administering a particular type of test to a certain pupil. This is likely to be the case if the counselor has already failed in an attempt to establish rapport with the pupil.

SELECT EXAMINERS CAREFULLY

There are several suggestions to assist examiners to do a good job. If more than thirty pupils are being tested under the supervision of one examiner, proctors should be provided to assist in the distribution and collection of test materials. In the lower elementary grades, proctors may be needed if more than ten pupils are being tested in a group. These assistants may be teachers or older pupils.

The person in charge of the testing program should meet with the examiners and proctors to study the test as well as the test manual. They should discuss the exact procedure to be followed and the problems or questions which are likely to arise. Examiners will profit by taking the test themselves. By so doing, they frequently become aware of difficulties

due to unusual positions of pages, typographical layout, or the method of indicating responses.

The following suggestions have been found helpful:

SUGGESTIONS FOR PLANNING TESTING PROGRAM

1. Provide examiners with written instructions supplementing and clarifying the test manual.
2. Prepare written instructions for all pupils if testing program necessitates major changes in the daily schedule. If testing is to be done during regular class periods with no major interruption of the daily schedule, no advance notice need be given pupils.
3. Avoid using pupil's free time or study time for testing.
4. Plan with administration so that administrative interruptions, e. g., fire drill, will not occur.
5. Avoid testing immediately after unusual physical exertion.
6. Schedule rest periods or recesses if several tests of considerable length are planned.
7. Test all groups at the same time, if several different groups are to take the same form of an achievement or aptitude test. This rule may be ignored with interest and personality tests.
8. Have all absentees take tests missed as soon as practicable.
9. Provide adequate working surface for easy manipulation of all test materials. Plain seats without desk arms are unsatisfactory. Individual desks are better than large tables or desk armchairs.

Good examiners will think of many preparations they can make in advance of testing. Everything they can do to assist their pupils to do well on the tests without violating the directions and intentions of the test manual will make their task easier. Some ways in which examiners can help make taking the test a more satisfying experience for their pupils are:

TIPS FOR EXAMINERS

1. Be particularly careful that the physical aspects of the testing room are good. Lighting, ventilation, heating, and freedom from unnecessary crowding are important.
2. If classes normally change during testing period, advise pupils in introductory statement to ignore the signal.
3. Have a supply of extra test materials, e. g., pencils, erasers, and scratch paper if needed.
4. Put a sign on the outside of the test room door to prevent unnecessary interruptions.
5. Use alternate seating, adequate proctoring, or other devices to encourage self-reliance during the testing, rather than warn against cheating.

6. Supplement oral instructions with blackboard illustrations for filling out basic data and other explanations allowed in the test manual.
7. Make sure pupils remove all extraneous books, clothing, etc., from the working surface.
8. Avoid arousing undue emotional tension by your own attitude or actions. Be matter-of-fact; the test is neither a crisis nor a lark.
9. Follow directions *exactly*, but don't be rigid and stilted in doing so. You can attain this goal by being familiar with the contents of the manual.
10. Make notes of individual atypical behavior during test. Anecdotal records of a pupil's behavior during a test are very important in interpreting his score. Observe also any significant reaction of a group or of the whole class for the same reason.
11. Collect test materials promptly and completely.

MAKE PLANS FOR SCORING

If we have done a high-grade professional job in administering our tests, we should not waste our efforts by a careless clerical job of scoring and recording the results. On the other hand, many tests have been adapted for scoring by means of a test-scoring machine. This machine is not sold, but is distributed on a rental basis.

Since many small schools would need the services of such a machine only a few hours a year, even rental is out of the question for one school alone. Several small schools in a county or school district might be able to coordinate their testing programs so that the cooperative rental of a test-scoring machine would be economically feasible. Many colleges have installed test-scoring machines and will score tests for schools at relatively small cost. Some publishers who sell tests adapted for machine scoring also are equipped to score tests. There are several agencies, such as the Educational Records Bureau, which include scoring among the other services they offer subscribers. Thus there are several possibilities to investigate in deciding how to have the test scored.

It may be that none of the suggestions for the machine scoring are feasible for a particular school. And it may be impossible to hire adult clerks. If so, scoring will have to be done either by members of the school staff or by pupil clerks.

SCORING SHOULD BE CHECKED

Clerical abilities of teachers are varied. The best examiner may be the poorest scorer. In general, the routine process of scoring objective tests has in itself little value for teachers. It is a chore to be done promptly and accurately. If teachers are utilized for the job, the school administration can make appropriate arrangements to minimize their working time

in scoring the tests. At least one member of the staff, preferably the counselor, should rescore the first few papers of each teacher to make sure that no one is systematically scoring incorrectly. An independent rescoring of every fifth answer sheet is essential to control the accuracy of results. If this audit reveals any one scorer to be consistently inaccurate, all papers scored by this person should be rescored.

Neatness and uniformity in checking responses and recording scores prevent errors and assist the auditor. All computations, particularly the addition of part scores as well as those involved in the scoring formula and in converting raw scores to norms, are sources of gross errors. They should be rechecked for each paper. This type of checking is especially valuable if the teacher scores the papers of pupils in his own classes. He is concerned only with the final results obtained. When he runs across scores which seem to him to be out of line with his knowledge of a pupil, he can immediately rescore the test. Even if he finds the paper scored correctly, he has focused his attention on an individual whose performance needs careful analysis. He and the counselor may profitably study all the relevant data in the pupil's record in the light of this discrepancy between test performance and the teacher's judgment. Of course, these discrepancies are regular causes of conferences between counselor and teacher regardless of the teacher's having scored his own pupil's tests.

Pupil clerks, too, can assist in scoring. Most of the practices suggested for teachers apply also when pupils aid in scoring. They must be mature individuals with interest, and preferably some training, in clerical work. The exploitation of pupils is not encouraged. If, however, the school and the pupils mutually accept a school work-experience program, test scoring is as defensible as other routine clerical tasks.

PUPILS MAY ASSIST IN SCORING

Some device to conceal the identity of the testee from the pupil scorers may be desired. A simple way to do this is to prepare in advance a numbered slip for each pupil. At the beginning of the testing period, these slips are distributed. The pupils are instructed to write their numbers on their answer sheets instead of their names, and to write their names on the numbered slips. The slips are collected so that a roster can be made listing each pupil's name along with his testing number. Without the roster, the pupil clerk will not be able to identify readily the individual whose paper he is scoring.

A RECORD OF TEST SCORES IS ESSENTIAL

The recording of test results in the cumulative record is also, in the main, a routine clerical task. What should be recorded? Four items are obviously essential:

1. Complete title of test and the form used.
2. Date administered.
3. Type of norms recorded, which must include:
   a. Kind of statistic, e. g., percentile or standardized score.
   b. Base population, e. g., local or national.
4. Raw scores and norm scores.

The first three items can be recorded by any competent clerical help available. Discretion is needed, however, in selecting personnel for recording the raw and norm scores.

Most of this information will have to be entered in abbreviated form on each cumulative record. For this reason, we shall find it valuable to maintain a separate file describing in considerable detail the test used, the conditions under which it was administered, the meaning of the norms recorded, analyses of group results, the raw score data, who administered and scored the test, and who recorded the results.

It is frequently wise to record several norms based on different groups. Eileen's standing on a scholastic aptitude test in relation to the other members of the senior class is useful in counseling her with regard to her present achievement. But her standing in relation to a large group of college freshman is more pertinent when going to college next year is discussed. Even though local norms are exceedingly valuable in many counseling situations, national norms should also be entered on the cumulative record. This practice enables us to make more accurate judgments about Eileen's probable success in various activities outside the local situation. It will also help the counselor at another school to which Eileen may transfer.

NORM SCORES SHOULD BE RECORDED

Many counselors have found that if they have a large number of test results recorded on a pupil's cumulative record, they have considerable difficulty in organizing these data in their minds. Significant relationships are confused by the presence of data which are relatively unimportant for the solution of a particular problem. The view of the forest is obscured by the trees. Test scores are sometimes recorded in the form of a statistical table on the cumulative record. Such a presentation is difficult for many counselors and teachers to interpret.

To overcome this difficulty, cumulative records sometimes provide space for entering the data in graphic form. Some schools have prepared special forms, tailored to their own testing programs, which are mimeographed so that each pupil's graphic testing record can be kept with his cumulative record. The individual profile charts which many publishers provide with batteries of tests from which several different scores are

PROFILE CHARTS ARE USEFUL

obtained are suggestive of the form which such records may take. Regardless of the exact form adopted, the record should provide for the accumulation of test data throughout the period covered by the general testing program.

Some non-test objective data, such as age, visual acuity, height, weight, number of siblings, and the like, can be converted to norm scores and also entered on the graphic record. It is desirable, however, to avoid cluttering the chart with information. The main purpose of the cumulative profile is to reveal quickly the high and low points, significant trends, and relationships. If we attempt to plot too many data, these major issues will be obscured.

There are several cautions to be remembered before a decision is made on the use of a graphic testing record. For many schools, not the least of these is the additional clerical work involved in plotting the norm scores and accurately drawing the graphs. We shall have to decide whether or not the advantages in ease of interpretation are worth the extra effort involved in the construction of cumulative profiles. And we shall need to recall everything said about the comparability of norms. Since the graphic record reveals at a glance the high and low points in Frank's testing record, it is doubly important that these points are not high and low, respectively, simply because his performance has been compared with groups to which he did not belong. Finally, we shall need to remember that important clues to the solution of many counseling problems are found among the non-test data in the cumulative record. The ease with which graphic test records may be interpreted should not lead us to neglect other available information.

Chapter V

USING TEST RESULTS

Our most important concern in modern education is for the individual pupil. We are concerned here with the use of test results for his benefit.

SCORES ARE RELATIVE— NOT ABSOLUTE

It is well to remember that test scores are not absolute numbers which represent a given amount, but, rather, they are numbers which indicate a relative condition. For example, consider a pupil with a high score on a scholastic aptitude test. We do not mean that he has 86 percent of a perfect ability to learn. When we say he ranks at the 86 percentile, we mean that he equals or exceeds 86 percent of the pupils included in the standardization group. This concept of the relativity of scores is basic to the rest of this chapter.

Now if scores are indicators of relative position or condition, our comparisons must all be relative. Does this mean that test scores are not exact? Does it mean that we have to lose our faith in the test results? Not at all. But it does mean that we cannot be too careful about the analysis of our test results. For example, a score of 43 does not mean the same for all pupils making that score. A boy of twelve may have much greater ability than one of sixteen, yet they obtain the same score on a test. The same number represents two different conditions. To get an estimate of the true meaning of test scores we translate them into percentiles or some other standardized score. These translated scores or norms indicate the relative position of pupils within the group with which they are compared. In the case cited, the boy of twelve might have a high percentile rank when compared to other twelve-year-old boys, and the boy of sixteen a low percentile rank. Of what value is the relative position of individuals? Can we make statements about individuals that will be of value? The answer is yes—If:

Yes—If we can be reasonably certain that our test is reliable and we will get essentially the same results each time we give the test.

Yes—If we have evidence that the test measures what we think it measures.

Yes—If the individual concerned is compared with a group to which he belongs.

If these conditions are met, then the test can be used successfully in helping individuals.

When we went to see the doctor last time, what information did we take along? Probably we were ready to tell him about the aches and pains that we were having. It is not likely that we were prepared to tell him the cause of our ill health. We may have had a theory about the cause, but we went to the doctor to get help. Before he could help us, he made a diagnosis. He set up, in his mind at least, a tentative working hypothesis. On the basis of this hypothesis or diagnosis, he began his treatment.

COUNSELORS MUST DIAGNOSE PUPIL PROBLEMS

Most of us are concerned with the problems of pupils. Sometimes pupils come to us with an accurate statement of the cause of their problems. Usually they are able to offer little more than a list of symptoms. They want help in finding out what causes these symptoms. Many times the counselor will need additional information. The doctor gets additional information by the use of questions, by observation, or from the results of clinical or laboratory tests. The counselor uses these techniques, too. He uses them to get a picture of the child. He uses them as a basis for making his diagnosis. When the diagnosis is made, he is ready to assist the pupil with his problem.

We have then these two points:

1. Scores are indicative of relative position or condition; they are not absolute amounts.

2. Counselors must make judgments if they are to assist pupils with their problems.

FOUR METHODS OF IDENTIFYING PUPIL PROBLEMS

How, then, can we help pupils to discover problems? One way is to use a scattergram. Darley uses the term "scatter-diagram" and the Germanes "quintile classification" to describe somewhat similar techniques. Exact references to these discussions are shown in Appendix A. For example, let us take the average of all marks made in the eighth grade by each pupil now enrolled in the ninth grade of Burchran Community School. There are 39 pupils in the ninth grade. Scores on the Henmon-Nelson Mental Ability Test, H. S. Examination given at the end of the first semester in the eighth grade are also available.

THE SCATTERGRAM

The marks were added together on the following plan: 4 for A, 3 for B, 2 for C, 1 for D, and 0 for E or F and divided by the total number of marks. This process gave the average mark for all subjects combined, shown in column 2 of Table 1.

The raw scores on the test have been converted to percentile ranks. For all practical purposes raw scores can be used instead of percentile

TABLE 1

PRESENTATION OF DATA FROM BURCHRAN COMMUNITY SCHOOL TO ILLUSTRATE METHODS OF IDENTIFYING PUPIL PROBLEMS.

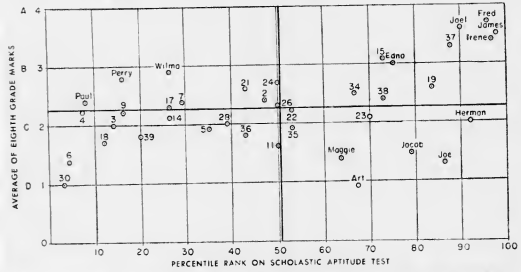
1	2	3	4	5	6	7
Pupil number and name (Given only for those discussed in this section)	Average of all 8th-grade marks	Score on Henmon-Nelson Test of Mental Ability, H. S. Examination	Percentile rank on mental ability test—8th-grade norms for Burchran Community School	Rank in class based on average marks shown in Column 2	Rank in class based on mental ability test scores shown in Column 3	Difference between average mark and mental ability ranks (Column 6 minus Column 5)
1. Edna	3.0	45	75	7	10	+ 3
2.	2.4	36	47	15.5	24	+ 8.5
3.	2.0	25	14	26	34	+ 8
4.	2.2	21	7	21	36	+15
5.	1.9	33	35	28.5	26	− 2.5
6.	1.4	17	4	35.5	38	+ 2.5
7.	2.4	31	29	15.5	27	+11.5
8. Herman	2.0	53	92	26	4	−22
9.	2.2	26	16	21	32.5	+11.5
10. James	3.5	61	98	3	1	− 2
11.	1.6	37	50	33	20	−13
12. Paul	2.4	30	8	15.5	37	+21.5
13. Maggie	1.4	41	64	35.5	16	−19.5
14.	2.1	22	26	23.5	29	+ 5.5
15.	3.1	44	73	6	11.5	+ 5.5
16. Perry	2.8	26	16	9	32.5	+23.5
17.	2.3	30	26	18.5	29	+10.5
18.	1.7	24	12	32	35	+ 3
19.	2.6	48	84	11.5	8	− 3.5
20. Irene	3.4	57	97	4	2	− 2
21.	2.6	35	43	11.5	22.5	+11
22. Jacob	1.5	46	79	34	9	−25
23.	2.1	43	70	23.5	13	−10.5
24.	2.7	37	50	10	20	+10
25. Fred	3.7	56	96	1	3	+ 2
26.	2.3	37	50	18.5	20	+ 1.5
27. Wilma	2.9	30	26	8	29	+21
28.	2.0	34	39	26	25	− 1
29. Joel	3.6	51	90	2	5	+ 3
30.	1.0	16	3	38	39	+ 1
31. Joe	1.3	49	86	37	7	−30
32.	2.2	38	53	21	17.5	− 2.5
33. Art	0.9	42	67	39	14.5	−24.5
34.	2.5	42	67	13	14.5	+ 1.5
35.	1.9	38	53	28.5	17.5	−11
36.	1.8	35	43	30.5	22.5	− 8
37.	3.3	50	88	5	6	+ 1
38.	2.4	44	73	15.5	11.5	− 4
39.	1.8	28	20	30.5	31	+ .5
Total	88.9					
Average	2.28					

ranks in the preparation of a scattergram. The percentile ranks are based on several years' testing of eighth-grade pupils in Burchran Community School. These data are recorded in columns 3 and 4, respectively, of Table 1.

To get some indication of the relative value of the marks, we like to know the average mark or, in the language of the statistician, the mean. We add all the marks and divide the total by the number of pupils. From Table 1 we add the figures shown in column 2 and get a total of 88.9. This, divided by 39 pupils, gives 2.28 which is the average mark for the pupils under consideration. We cannot do the same thing with the percentile ranks because percentiles are not equal units throughout the range. Therefore, we take the point which separates the upper half from the lower half. This point is known as the median. In this illustration, the percentile rank of the twentieth person divides the group into halves. Simply by crossing out the 19 lowest scores in column 4 of Table 1, we find the median percentile rank to be 50.

Now let us look at the scattergram. The average marks and percentile ranks from Table 1 are portrayed graphically on the scattergram. Consider

SCATTERGRAM OF NINTH GRADE PUPILS*



*Based on data shown in Table 1. Names shown for pupils discussed in text. Other pupils identified by number.

the data given for the first pupil, Edna. She ranks at the seventy-fifth percentile on the Henmon-Nelson test. The percentile ranks are graphed according to the scale at the bottom of the scattergram. On this scale we move from left to right until we come to 75 (between 70 and 80). This is her position on the percentile scale. Her marks averaged 3.0. Marks are graphed according to the scale at the left of the graph. Thus we follow an

imaginary perpendicular line at 75 from top to bottom until this line intersects with an imaginary horizontal line coming from 3.0. Where the two lines intersect, a dot is placed. This point represents the average mark and percentile rank of Edna. In like manner the scores of the other pupils are recorded on the scattergram.

After the scores and marks for each pupil have been recorded on the scattergram, we draw a line across the scattergram at 2.28. This indicates the average mark earned by all pupils. All pupils above this line have marks above average; pupils below the line have marks below average. Next we draw a perpendicular line at 50, the median percentile rank. All pupils to the left have scores in the lower half of the group, whereas pupils to the right have scores in the upper half.

We consider both of these lines at the same time. They divide the scattergram into four sections or quadrants. The pupils in each of these quadrants can be described as a group. The description for the upper right quadrant is above average in ability and above average in marks. The lower left quarter can be described as below average in ability and below average in marks.

The pupils in the upper left quadrant can be described as below average in ability, but above average in marks. The lower right section contains those pupils above average in ability and below average in marks.

Thus we have four general types of persons which make up the scattergram, those with:

1. Low ability and low marks.
2. High ability and high marks.
3. Low ability and marks higher than expected (the overachievers).
4. High ability and marks lower than expected (the underachievers).

The scattergram, therefore, makes it easy to identify each of these types of pupils.

There are certain limitations in this technique. Germane and Germane discuss the following in their book.

1. It is difficult to get an accurate measure of ability to learn.
2. Teachers' marks are frequently not reliable or valid measures of achievement.
3. Achievement tests, standardized or teacher-made, may not be an accurate measure of achievement.
4. High ability pupils can make a high score on a subject-matter test by cramming. Little real learning takes place although the scattergram would place them in the upper right quadrant.
5. The scattergram tends to focus attention on scholastic achievement.

If the school is too subject-matter-centered, the scattergram might motivate a drive for increased memorization. Some teachers might use it exclusively as a device for prodding the underachiever.[1]

TEST PROFILE

The cumulative record of the individual pupil frequently contains a profile or chart. On a test of scholastic aptitude Gerald ranked at the seventy-eighth percentile of sixth-grade norms. On a test of general achievement, he ranked at the twenty-fifth percentile on sixth-grade norms. Obviously, we could reach no other conclusion but that Gerald was underachieving. A rank in the highest quarter on ability as opposed to a rank in the lowest quarter on achievement is a certain indication of underachievement if other factors such as test reliability and validity are satisfactory. If we took Gerald's score and plotted it on the scattergram, it would fall in the lower right quadrant. Although we have compared him against a whole group, we are basing the comparison on the discrepancy between his scores.

COMPARATIVE RANK IN GROUP

A third method of classifying or identifying the pupils who fall in each of the four categories is to compare ranks. For example, let us use the list of pupils in Table 1. Fred has the highest grades in the class so we put a *1* after his name. Joel has the second highest grades so a *2* is placed after his name as shown in column 5 of Table 1. On the scholastic aptitude test James has the highest percentile rank, so a *1* is placed after his name in column 6. Irene has the second highest percentile rank so a *2* is placed after her name in the same column. When the column 5 is subtracted from column 6, we obtain the figures as shown in column 7. Those with a minus sign indicate that ability is greater than achievement or, in other words, they are our old friends, the underachievers. Those with large discrepancies are the more pronounced cases; certainly minor deviations in either direction are not significant.

ACCOMPLISHMENT QUOTIENTS

For a number of years it was common practice for schools, particularly at the elementary level, to compute Accomplishment Quotients. These were simply the Educational Age divided by the Mental Age. The resulting quotient was interpreted in much the same manner as an I. Q. A score of 100 meant that the pupil was achieving at the level expected; 120 meant he was overachieving; and 80 or below, he was underachieving. These quotients have fallen into general disuse because the Educational Age and Mental Age were seldom based on the same sample. Consequently, the A. Q. had little real meaning. This same criticism applies

[1] C. E. and E. C. Germane, *Personnel Work in High School* (New York: Silver Burdett Co., 1941), pp. 110-14.

to methods just described unless local norms are used. In addition, the work involved in the computation of A. Q. was found to be disproportionate to the value of the results obtained.

Thus the four following methods of relating pupils' achievement to ability have been discussed: (1) scattergrams; (2) profiles; (3) comparative ranks; (4) Accomplishment Quotients. Since these methods have certain basic assumptions in common, they are noted here.

First, the measure of achievement is assumed to be reliable (consistent) and valid. As a matter of fact, measures of achievement are not perfectly reliable nor always valid. Consequently, we must expect some error. It can be expected, for example, that Joel's raw score on an achievement test may vary 5 or 10 points. In constructing a scattergram, this might well put Joel in the underachievement group when actually he should be in the normal group. The same kind of an error can occur because the test does not measure achievement only. In other words, it is not 100 percent valid. A test in chemistry may require considerable reading ability so that poor readers make poor scores, not because they do not know chemistry, but because they cannot read fast enough to get the problems done in the time allotted. In such a case, classifying the pupil as an underachiever in chemistry is inaccurate. These two sources of error must be considered constantly when it comes to the interpretation of scattergrams.

Measures of ability are affected by the same source of error as measures of achievement. Thus we have to consider both our measures as approximates rather than absolutes. One way to think of it is to consider the dot on the scattergram as the center of a circle. The lower the reliability and validity of either the measure of ability or achievement, the larger the circle. Somewhere within the circle lies the true score, but the circle may be so large that it covers half the scattergram. With even the best of tests, the circle has a diameter so great that we can hardly trust the identification of students that fall near the boundary lines of our quadrants.

Another basic concept is that it is easier to get a score lower than the true score than it is to get one higher. Frank can actually spell 80 percent of the sixth-grade words. He takes a spelling test from a teacher whose pronunciation is indistinct; he rates 76 percent on the test. There are numerous other reasons for Frank's getting a lower score than he deserves. Rarely will he take a spelling test and be able to spell 86 percent of the words. Consequently, we can usually be on safe ground if we interpret scores as a slight underestimate of achievement.

Still another basic concept is that both extremely high or extremely low scores in any test are rare, and pupils who get such unusual scores in one test are likely to get scores closer to the average on a second test of a

similar nature. If Wendell makes an extremely high score on a scholastic aptitude test, we might expect him to make an exceptional score on an achievement test. The odds are that his score on the latter will be above average, but it is quite probable that this second score will not be as high as the first. In fact, had he taken another scholastic test, the chances of his bettering or even matching his first rare performance are not nearly as good as are his chances of getting a lower score. We must remember this when we try to explain the discrepancies between exceptional ability and achievement. When one score is extremely high or low, the tendency for a second score is to be nearer the mean.

USING THE SCATTERGRAM

Now let us look at the pupils in the upper left quarter. The majority of scores falling in this category are those of pupils whom we call over-achievers. They are apparently achieving at a higher level than is expected for their ability. We can be reasonably certain that, if we actually have an over-achiever, we can explain his behavior in terms of excessive motivation or excellence of study habits. First, however, we must check to be certain that a reading deficiency did not bring down his mark on a timed scholastic aptitude test so that the low score makes him look like an overachiever. Other errors in testing can occur, such as the tendency for an extreme score to be paired with a score not so extreme. *If we are reasonably certain that our measures of ability and achievement are accurate, then we can begin the process of determining the cause of the discrepancy.*

CHECK MEASURES OF ACHIEVEMENT AND APTITUDE

A special word of caution: If teachers' marks are used as the measure of achievement, they may be biased. Wilma is an attractive girl and quiet in class. She hands her work in on time and never has caused a disciplinary problem in class. She, from the teacher's angle, is a nice pupil and is given a grade of B. When this grade is compared with her ability, it appears that she is an overachiever. Actually, instead of being an overachiever, she is overrated. When Wilma comes up against a teacher who marks only on the basis of achievement test results, she may be in academic difficulties. Many of these overrated pupils do very well in the occupational world where their pleasant personality is a definite asset. We must remember, though, that eventually they will have difficulties in school unless they select subjects in which, and teachers with whom, they can succeed. This does not mean that all pupils with low scholastic ability should be assigned to the shop or home economics because they can get by in those courses. Quite to the contrary, they should be assisted to select courses in which they have a genuine interest and in which their special abilities, even if somewhat limited, can be used to the full. These pupils who have tasted academic

success (at least obtained good marks) must be prepared for the day when their personality will not see them through.

Now let us assume that we have eliminated the erroneous instances by checking for errors of measurement. We have sought to isolate those who are overrated rather than overachievers when marks are used as our criteria of achievement. When the diagnosis is overrated, we can easily check our hypothesis by administering an achievement test which provides a more or less objective measure of achievement. After this has been done, there still remain a few scores in the upper left quadrant that are not explained. Perry is a typical overachiever. What is the reason for his zeal? His father, an educator, may constantly tell him how important it is that he be successful in school. The test results have helped us locate him; now, what can be done to assist Perry? The first and obvious thing we shall want to do is to check Perry's mental and physical health. If no nurse or doctor is available, it is possible for us to make some simple observations regarding his health. Does he stutter? Is he underweight? Does he take part in a normal number of activities or does he devote his full time to study? We continue asking questions of ourselves or Perry until we come to a conclusion about Perry's present mental and physical health. Suppose that we decide that it is good. Does that end the matter? No! We must begin at once to build for the future adjustment of Perry. At present, he is satisfied to devote a disproportionate amount of his time to study. By devoting this extra time, he is able to achieve satisfactorily even though he spends most of his time studying. Said another way, Perry will probably reach the limit of the amount he can learn regardless of his extra effort or the attempts made by his teachers to adapt their instruction to his limited ability. When he reaches this point, he should be prepared to accept some compromise.

THE OVER-ACHIEVER

The preparation for this compromise is the important task of the counselor. In essence, the counselor's job seems to be, not to tell Perry that he will fail, but to help him widen his interests and develop attitudes which will make his future adjustment easier and more satisfactory. Perhaps he should be encouraged to take a more active part in the extra curricular activities of the school. Maybe his excessive studying has shut him off from social contacts. If so, Perry should be encouraged to include more social functions in his schedule. However, of equal importance is the matter of developing attitudes. At present he is apparently convinced that the most important thing that exists for him is to maintain a high record of scholarship. Admirable as his attitude may seem to us as teachers, it is not without danger.

Paul, another overachiever, differs from Perry in that he is not in

good mental and physical health. The attitudes necessary for Perry to develop as a protective measure for future adjustment are not present in Paul's makeup to support his present adjustment. Although his grades in general are above average, he is failing algebra. He is unable to comprehend it. He studied harder and still failed. This cycle has kept up until now he is cross and irritable. He has the idea that good grades are the most important thing in life. He refuses to take part in co-curricular activities because he needs the time to study. How did he get this way? What made him so ambitious? It is a long story and considerable counseling skill was required to obtain the facts. It can be summarized by saying that his mother feels she married below her station. She cannot, she thinks, be proud of her husband. She is determined that Paul do something of which she can be proud. By pressures, such as cash rewards for A marks or restriction of chances to go to the movies on Saturday for low ones, she has built in Paul a terrific drive for good marks. Now that he has reached his limit, he can profit little from continuing in a formal education. He has few other interests. His drive in school has been marks rather than knowledge. He is not aware of his limitations. He is unhappy and may soon become a disciplinary problem; he may become morose or he may drop out of school.

We are clearly faced with a serious problem. If we do not think we can handle the treatment of a personal adjustment problem successfully, then we are responsible for referring him to some person or agency which can help him. It is not enough to send Paul to see someone. We must give all the information we have about Paul to the person who is undertaking treatment. In addition, we must make a careful follow-up of such referrals to determine, first, if Paul arrived and, second, what we can do to assist in his adjustment process.

But whether we refer a pupil to someone else or try to assist him ourselves, we should be aware of the causes of overachievement. In essence, excessive drive can be caused by pressures of diverse character. Among them are: Parents overemphasis on study, pupils' desire for recognition, fear of failure, and unduly limited interests. These pressures must be relieved or diverted before successful adjustment can take place. This can be done in a number of ways. Parents can be convinced that they should set more appropriate goals for their child. The pupil can, by being informed, accept substitute goals, such as striving for success in a co-curricular activity in line with his abilities rather than trying to master trigonometry. He may be helped by learning of his abilities and what level of achievement he can expect. In summary, the treatment boils down to two phases: First, the identification of the pressure and its cause; second, the relief or redirection of that pressure.

From this discussion we might conclude that all overachievement is bad because of tensions created in the individual. But if we review our experience in the classroom, we can recall a good many pupils who achieved above measured ability because of good study habits and consistent, methodical work. There was no tension. They simply knew better than most of their classmates how to attack a task and carry it to completion. Not every overachiever is the unfortunate victim of a dominating father or a neurotic mother! Further, overachievement is identified by comparison of grades in specific subjects with results of scholastic aptitude tests, most of which are omnibus measures. A pupil may overachieve in certain areas simply because he had high aptitude for work in those areas—aptitude which is obscured by the nondiagnostic character of the measures of scholastic aptitude.

Now we come to the pupils who fall in the lower right quadrant of the scattergram. These pupils are in the upper half in scholastic aptitudes, but they are achieving below the average of their fellows.

THE UNDERACHIEVER A term commonly used to describe this group of pupils who are achieving less than expected is underachievers. Just as we found explanations for the behavior of overachievers, we can locate the causes of underachievement. One explanation for underachievement is that the pupil is just lazy; but this is probably a poor explanation. It is quite likely that most persons underachieve for other reasons.

Let us examine some of the causes. As with overachievers, before we make any other investigation of discrepancies, we check our measures of achievement and ability. Retest if possible; use other judgment-making devices such as rating scales, teacher opinions, and other evidences of achievement or ability which may be gleaned from the cumulative record. After this check has been made and we are certain that a real difference exists between achievement and ability, then we begin the task of discovering the reasons. What then are some of these reasons? An easy one to check is the attendance record. If an excessive number of absences or an extended absence appears on the record, we have a lead to follow. Marian cannot achieve if she is not in school. If this should be the only explanation of the low achievement, the problem may be solved by providing make-up instruction for her. However, a more fundamental approach is to discover the reason for her excessive absence. At any rate, Marian's case is not one of underachievement, but lack of opportunities to learn.

Another cause of underachievement is lack in one of the fundamental skills. Particularly is this true of reading ability. For almost all subjects, the ability to grasp meaning from print is basic. What does this mean in terms of individual counseling? For all cases of underachievement, it is at least necessary to consider the possibility of a reading problem. Be-

fore giving extensive tests, it is best to make a rough check. We can have the student read silently a few pages of new material, time him, and then ask questions about the article. One counselor kept a three-page travel booklet on his desk and used it with all his pupils. After a while, he had a rough idea of the length of time required to read it. He was then able to make judgments about the speed of reading of his underachievers. The level of comprehension was checked in the same manner. By watching for frowns, squints, lip movement, and so forth, as the pupil read, he was sometimes able to make good guesses as to the kind of reading difficulty. If any indications were found that a deficiency did exist, then he followed with a more complete diagnostic examination with standardized reading tests.

DEFICIENCY IN BASIC SKILLS CAN CAUSE UNDER-ACHIEVEMENT

It is not only a lack of reading skill, however, that affects achievement. Lack of skill in arithmetic easily causes failure in algebra. Lack of basic skills in grammar and punctuation has caused apparent underachievement in advanced English.

It is essential that something happen besides the identification of these deficiencies. Remedial instruction must be undertaken. One of the functions of the guidance program is to provide information for the faculty and administration. The counselor should constantly try to provide the information that the faculty and administration need to discharge their responsibility for remedial instruction.

In addition to each of the skills specifically related to subject matter, a common cause of underachievement is poor study habits. Study activities, according to Traxler can be put into two classifications:

STUDY HABITS AFFECT ACHIEVEMENT

1. *Work habits* which include "keeping in good physical condition, planning a definite schedule, forming regular habits in regard to time and place of study, avoiding distractions, getting clearly in mind just what is to be done."

2. *Study skills* which are such activities as "note taking, outlining, using books and the library, or problem-solving."[2]

Traxler believes that classroom teachers can assume much of the responsibility for training in study habits and skills. The first step in helping a student acquire good study habits is a thorough case study. This study should be made by the counselor with the help of other teachers. It frequently is found that the counselor can help solve the problem by discussing a study schedule or time budget with the student. But if it is a problem involving a study skill, then instruction in that skill is the only way to

[2]A. E. Traxler, *The Teaching of Corrective Reading in the Junior and Senior High Schools* (Bloomington, Illinois: Public School Publishing Co., 1942), p. 24.

help. The first step is to find the specific kind of study deficiency. The second is to make plans for eliminating it. The third is to follow-up the plans to ascertain that the skill is being acquired.

What cause other than study habits can we find for failure to achieve up to the expected level? Out-of-school work at times interferes with achievement. Joe works so many hours that he comes to school tired out. Maggie is so interested in her outside work that she neglects her school work. Herman would like a job which did not interfere with his school work, but cannot find one that pays enough for him to keep on helping to support his mother. And Art does not like school, but keeps on to satisfy his folks. He works because he can be successful at it, and he is always failing in school. These four are typical of the kinds of pupil problems that will be found in this area. Solutions to them usually follow the same pattern. The pupil is placed in a new job that is more in line with the hours he has available for work, or his school schedule is adjusted to fit his needs.

These pupils comprise a large proportion of drop-outs from school. The prevention of a premature withdrawal from school is a function of the guidance program. How to keep Maggie with her other interests, or Joe with his repeated failure, in school tries our counseling skill. Possibly we shall be able to get them to achieve according to expectation. Rearrangement of the schedule, changing teachers, or getting a new job may be the answer. The really progressive school will devise work more suited to a pupil's needs. Or it may well be that we shall have to settle for underachievement and be happy about it.

Now we come to the last of the causes of underachievement that we shall discuss: the area of personal adjustment. A great many of us know that at times our own personal adjustment is poor. We lack the drive or motivation to buckle down and do the job. Or we become so concerned over our kid brother's latest exploit that we cannot concentrate. Sometimes the principal begins to pick on us, or our fellow teachers irk us. When these things happen, our efficiency for the day drops. And the more we fret about the thing, the less efficient we become. Just as these personal adjustment problems throw us off our stride, they bother boys and girls in school.

ADJUSTMENT PROBLEMS INFLUENCE ACHIEVEMENT

Instead of being annoyed with his kid brother's actions, George may be embarrassed by his parents' behavior. We have our iron-clad rules handed down to irritate us, but don't our pupils have theirs too? If we do not get along with all our fellow teachers, can we expect that Tom, Dick, and Harry will all be well adjusted to each other? No! We can expect that we shall always have our problems and that our pupils will have theirs.

If that is true, what is the purpose of considering the problems? Just this: problems of a temporary nature are normal, natural, and expected, but when the problems persist, then it is time for us to be concerned. We have little cause to worry if today Janet shows a strong dislike to Miss Wood, her teacher. But if Janet persists in her dislike, counselor beware! Janet may show her resentment by not studying for Miss Wood's class. She may become a behavior problem in class, begin to spread malicious rumors, or she may take the attitude that she will "show that old stick that I know more about the lesson than she does." Whatever attitude she adopts, she is not heading for a satisfactory adjustment if the basic drive comes from a hate for Miss Wood.

We can recall from our experience many other examples of poor adjustment. Ordinarily, the adjustment problem causing underachievement falls in one of these four categories: (1) lack of motivation, (2) home difficulties, (3) personal maladjustment, and (4) poor pupil-teacher relationships. Certainly we should look for evidences of maladjustment in these areas when we deal with underachievers.

So far, we have discussed the pupils who are not achieving at the expected level. Now we come to the largest group of pupils, the ones that are achieving at the expected level. These are found in the lower left and upper right quadrants of the scattergram. Even though all are achieving at the expected level, they may be divided into two groups: low ability and low achievement in the lower left section, and high ability and high achievement in the upper right square. Because they are quite different, let us consider each group separately.

LOW ABILITY AND LOW ACHIEVEMENT

What more can we say about low groups than that they are low? Aren't these the least-talented pupils in the school? Should we not pride ourselves that at least they are achieving up to capacity? Are they not normal, well-adjusted boys and girls in the school?

First of all, let us examine the measure of ability that we have used. Is it one in which the verbal factors of intelligence are measured to the exclusion of others? Can we expect that our measure of achievement is related to the measure of ability? For example, can we expect that a measure of achievement which includes grades in shop, art, and other non-verbal subjects is closely related to a verbal intelligence test? In all probability a person who does not possess verbal ability will appear low on the chart because most achievement measures are heavily weighted with the verbal factor. But what of the boy in this group who possesses talent in music or art, or the girl who has superior dexterity or other non-verbal characteristics. Are they going to be overlooked because they appear to be achieving up to their ability? We must provide some method of checking on their special abilities.

KEEPING BOYS AND GIRLS IN SCHOOL

The low boys and girls are the ones who are getting the majority of the low and failing grades in the school. They seldom taste success. They have every reason to become discouraged. Discouragement often leads to but one thing—dropping out of school. The youth who lacks ability certainly cannot be expected to continue in school if he has no accomplishment other than low grades. The low grades that he receives are probably indicative of the little that he has learned. If school is not vital to him, if he is not taught something that he sees has value, if he is constantly working in an area in which he knows nothing but failure, we can safely bet that he will drop out of school the first chance he gets.

We need to identify these potential drop-outs, but we need to do more than locate them. We must do something that will help them stay in school longer. This does not mean that we have to coddle them or that our academic standards have to be lowered. Rather it means that we have to make some provision for the instruction of these individuals in areas from which they can profit. The problem is not going to be solved by the introduction of a vocational training program in the school. The low-ability student may not be any more successful in shop work than in academic courses of the school.

For the pupils in the lower left section of the scattergram, is it not better to have them take some kind of vocational training than to try to continue in the academic world? The answer to this question is a definite No! It is estimated that about 50 percent of jobs do not require special training prior to employment. These routine jobs are the ones that are likely to be held by those in the low group. If they are to hold these routine and, to many of us, uninteresting jobs, is it not essential that the school provide some kind of training that will help them to a full life? It is at this point that the counselor's job takes on real meaning. His job is helping these pupils find activities within their capacity which will be of continuing value. Is it not good counseling to call attention to leisure-time activities that can be learned in the school that will make life more interesting after these pupils leave the school? Is it out of the scope of good education to teach these pupils to play, to derive benefit from recreation, to spend their leisure time doing something for themselves rather than spending a good part of it at commercial entertainment? Many counselors believe that one of the most effective uses that can be made of tests is the identification of these marginal pupils. After they are identified, the test results may be used to help them obtain an educational experience that will have some meaning for them, both now and in adult life.

If, then, we identify these potential drop-outs, we can assist them in two ways: to plan a meaningful school experience and to continue in school without the bugaboo of failure hanging over their heads.

In planning meaningful school experience, we must remember that the division of responsibility for identifying and effecting needed changes in the curriculum is clear-cut. The guidace program serves in an advisory capacity. It collects information which can be used as a basis for curriculum revision. It presents this material with recommendations to the administrative officials of the school. Beyond that point, the school staff as a unit takes over. Members of the teaching staff, pupils, and parents can also make recommendations regarding the school's offerings. But the final responsibility for organizing the curriculum rests with the school's administrative officers, assuming, of course, a democratic procedure of staff cooperation.

Now let us consider the second group of pupils who are achieving at the expected level, the ones in the upper right of the scattergram. These pupils have high ability and high achievement. At least, that is one theory that explains their presence in this section of the scattergram. They do have high achievement in terms of their fellows, but we cannot neglect considering the level of their achievement in terms of themselves. A few pupils find that when their high ability is contrasted with school requirements which are geared to the average student, it is relatively simple to be classed as a high achiever. They are able to loaf through the class work, spending only a little time cramming for examinations. Or they retain the information that they hear discussed in class. Their achievement is good. It is better than the majority of the pupils in the group. But it is not all that they are capable of doing. Because they are not required to work up to capacity, they may easily develop careless habits of study. They may fall into the assumption that school's a breeze. Here again, the counselor can make good use of test results. After these high ability pupils are located, he can help them understand that they have ability to do exceptional work. He can help them choose classes that will stimulate and challenge them. He can, through interviews, help them to determine proper goals and to make plans to reach them. It is this group of pupils which contains potential leaders. Because they are the probable leaders of tomorrow, extra effort on the part of the counselor to help obtain the maximum benefits from education would seem justified.

HIGH ABILITY HIGH ACHIEVEMENT

USING THE RESULTS OF INTEREST TESTS

Let us review the factors which we considered in our previous discussion of interest tests.

FIVE FACTORS WHICH AFFECT THE USE OF INTEREST TESTS

1. *Pupils' estimates of their job interest are not dependable.* Factors such as overestimating the earnings of an occupation, parental pressures, a drive for social prestige, or lack of occupational information may influence pupils' estimates unduly. The basic interests which make for real job satisfaction are frequently overlooked by a pupil as he estimates his interest in a specific occupation.

2. *Frequently, interests are not related to aptitude or ability.* An example sometimes used to illustrate this point is the soprano in a local choir who cannot sing but insists on punishing the congregation because she is interested in singing.

3. *Basic interest patterns are reasonably stable.* We all know of pupils that have changed their occupational plans dozens of times. Careful analysis of these plans will usually disclose that most of the occupations are closely related. Changing from electrical engineering to chemical engineering to mechanical engineering and, finally, to machinist are not drastic changes. Underlying all of these occupations is a basic interest in things and in technical matters. Change from one occupational choice to another may well be a process of seeking an occupational level which is interesting, or more accurately, satisfying, to the individual, or nearer his ability range.

It is true that specific interests will change frequently, but the research evidence seems to indicate that the basic pattern of interests is reasonably stable. It is this pattern which interest tests purport to measure.

4. *Interest tests are quite reliable.* That is, the results you obtain from one administration will be about the same as from another administration under the same conditions. They are about as reliable as scholastic aptitude tests.

5. *The validity of interest tests is a confused issue.* The authors of interest tests have not agreed on a method for determing the validity. One author has based norms for his test on the similarity of pupils' responses to those of persons engaged in various occupations. Another has attempted to isolate the responses which, when combined, yield a single score for an area of interest. And another has set up theoretical interest areas and built items which are thought to be indicative of interest of the specified type. The diversity of these measures has caused some counselors to be extremely skeptical of interest test results. To make matters even more confusing, available interest tests do not all yield scores in the same areas. For example, one test may have an agricultural score while the next one does not.

With these facts in mind, we may logically ask, of what use are

interest tests in the guidance program? Obviously, they must be used with unusual caution because research is not yet available to support their use as definitive instruments.

INTEREST TESTS USED FOR MOTIVATION

Many counselors usually find it difficult to interest students in occupational or educational planning. They have found, however, that the administration of an interest inventory has frequently started pupils thinking about future plans. Because interest tests do not have the connotation of being a *test,* pupils usually enjoy taking them. After the pupil has taken the test, he is interested in finding how he came out. Certain interest tests are designed so that pupils can score and prepare profiles for their own tests. These have the advantage of being more economical of time and money. They do have an inherent danger however in that pupils try to interpret their profiles unaided. The interpretation of interest tests is a technical process. It cannot be done by pupils or for that matter by teachers who are not trained in interest measurement.

INTEREST TESTS FURNISH CLUES FOR THE COUNSELING INTERVIEW

Interest test results help get an interview under way. The counselor can look over the responses to interest test items in his pre-interview study of the pupil. This will give valuable clues. If the pupil indicates that he would rather go fishing than play baseball and would rather hunt than play golf, the counselor has a clue he can use as an ice breaker for the interview.

Although this use of interest tests does not preclude the use of total scores, interpretations of total scores must be supported by research. For example, one widely used interest test yields total scores in several areas, for each of which the manual lists a number of occupations. To date, however, evidence is lacking that a high score in any particular area indicates interest in the specific occupations listed. Therefore, in using this test with a pupil, we can safely say only, "On the basis of this test, you seem to have greatest interest in this area." It is wrong to say, "Because you have highest interest in this area, you will be interested in these occupations." We have no evidence for that statement. We might say, "Your interests seem to be in this area. Many people who have interests similar to yours are believed to be interested in these occupations. Perhaps it would be well to investigate these and see if you could be satisfied with any of them." What then is the use of the total score on this type of test? *It furnishes a clue to a basic interest area;* it does not indicate interest in a specific occupation.

These basic interest areas may have little specific occupational significance in many cases. The number of persons, for example, employed in such interest areas as art or music is comparatively small. Because the

employment opportunities are limited, the competition is likely to be keen. With some pupils, we would probably confine our interpretations to: (1) avocational activities, such as hobbies, extracurricular participation, or other types of recreation which will give expression to the interest or (2) related occupations which provide for partial expression of interest such as a salesman in a music store or cataloger in an art museum.

DISCREPANCY BETWEEN MEASURED AND CLAIMED INTERESTS

Frequently, we find that the interest scores do not agree with the pupil's statement of his job interest. This should not cause undue alarm because research has shown that the pupil's statement of interest (claimed interests) are: (1) very transitory, (2) easily influenced by extraneous factors, and (3) often result from misinformation.

As we look back over our own experience we can remember many different occupations in which we claimed to be interested. Almost all children change their claimed job interest several times during their school experience. This is natural and is healthy, if the change is brought about by greater insight, new appreciations, or better understandings. The interest test results can be used to help pupils formulate plans for study prior to these changes. This is done by accepting the pupil's claimed interest as bona fide. The counselor should assure the student that the choice is his, and that the interest test results may be wrong. Some counselors say, "This test indicates that your interests may lie in a field different from the one you have chosen. Do not accept this at face value. But you should not completely disregard the findings either. Perhaps it would be well to consider carefully your present choice in view of the apparent conflict. How can this best be done?" Then they lead the pupil to see that it is his responsibility to seek additional information upon which to base a decision. Usually this involves getting more occupational information, especially that which deals with specific duties, and, if possible, try-out experiences.

In counseling with pupils having this discrepancy, the cause can frequently be traced to some outside factor. A maiden aunt who is a nurse and willing to pay the cost of a medical education may cause a boy to claim an interest in medicine. The policeman father who always wanted to be a lawyer may influence his son's choice of law as an occupational goal. But these factors do not always direct Jack or Jill toward an occupation. They are equally potent in steering away from any goal. The brother that failed as a businessman may discourage a similar choice. The dry science teacher may anesthetize a genuine interest in chemistry so that a substitute goal is selected. We must help the pupil to scrutinize carefully all factors which are contributing to inconsistencies between measured and claimed interests.

Thirdly, we shall want to consider with the pupil the possibility that his stated interests, or those revealed by the test, are based upon incomplete information or downright misinformation. Cromwell in an unpublished study supplied a group of pupils with a list of 100 occupations and asked them to put a plus sign in front of the ones that interested them, a minus sign in front of those they would not like, and a zero in front of those to which they had no reaction. Those terms which they did not understand were to be left blank. He included in the list two fictitious occupations, "Medical roustabout" and "Naval scavenger." Over 80 percent of the pupils were *interested* in each of these occupations. We commonly find that boys interested in engineering actually mean mechanics or that girls interested in missionary work think only of the travel and the romance of foreign lands. Sometimes the supposed high salaries of certain occupations or glamorized working conditions account for claimed interests. An insidious kind of misinformation results from the differences in grading standards among teachers. An easy teacher may give A for mediocre work while another may give B for superior performance. From the difference in grades, pupils infer that they are doing better in the A subject than in the B subject. From this inference, it seems logical for them to take the next step—"if I can do A work in that subject, I should choose that as my life work because I am more successful at it." Counselors need an intimate knowledge of the school, if they are to cope with this problem.

Unequal grading standards present a constant problem. As we search for evidence with which to resolve this problem, we may find that the crux is the discrepancy between interest, either measured or claimed, and ability. It is generally believed that we do those things well which are interesting to us, and like those things which we do well. Which comes first, ability or interest, is similar to the time-honored chicken and egg debate. We can in many cases accept as bona fide a relationship between interest and ability. The occurrence of a discrepancy between the two is frequent enough, however, to warrant re-consideration.

DISCREPANCY BETWEEN INTEREST AND ABILITY

Perhaps most of the causes could again be grouped into the three sections dealt with above, namely: (1) instability of interests, (2) extraneous influences, and (3) inadequate or incorrect information. Let us consider briefly the case of Herbert. On the personal data sheet he had indicated a desire to be a doctor. His scores on the interest test indicate that his most pronounced interests are in the scientific area. The manual for this test lists physicians as one of the occupations in the scientific group. There is apparently no conflict between measured and claimed interests. But before Herbert makes a final choice, he should consider the relationship of his interests to his ability. His grades in general science and biology

were average, but his present grades in chemistry are well below average. On the California Test of Mental Maturity he ranked at the 35 percentile of tenth-grade students in his school. This additional information raises the question, can Herbert meet the entrance requirements for medical school? The evidence would suggest that the counselor should lead Herbert to re-evaluate his choice of medicine. We have here a case of discrepancy between interest and ability.

How can we find the solution to it? First, let us examine the stability of the interest. During the interview we find that Herbert is somewhat upset about his failing grades in chemistry and he says, "I have been thinking about shifting over to physics or astronomy. I got along better in those parts of general science. That way I can get out of taking chemistry. Don't you think I would be just as happy in those sciences as in medicine?" From this conversation, we secure our leads for additional questions such as: how long have you been interested in medicine? What other occupations have you thought about? Why do you want to be a doctor? This would help us to determine whether or not the decision was one of long standing. The length of time that the pupil has been considering his choice influences our methods of counseling. Usually, the longer the pupil has held to a choice, the greater his emotional reaction to any change of plans.

In cases like Herbert's, the second thing to determine is the presence of outside factors which are influencing his choice. If we find them operating, the problem may be one of conflict between interest and outside pressures rather than interest versus ability.

The third area, that of information, is ordinarily the most crucial. An investigation of the experiences which Herbert had in general science revealed these facts: He liked to read craft magazines; from one of these he took the plans for making a telescope; as time went on, he gained a rudimentary understanding of the principles of optics; he became acquainted with other amateur astronomers and even ground some lenses in his workshop. This scientific interest was largely confined to applied aspects, particularly the craftsman jobs. The telescopes he built were finished products but his use of them was limited. He enjoyed doing things and thought this required scientific training. When he faced again the cold realities of scientific theories, he lost his interest. He lacked adequate information about the opportunities for the type of work he liked. He was misinformed about the duties of some occupations and was dealing with job titles rather than duties, responsibilities, and rewards of the occupation. After he was helped to secure adequate information, he selected optical worker as his job goal. After an apprenticeship with the local optical manufacturer, he became a successful and happy lens grinder.

There is danger that many counselors place too much reliance on a single isolated interest test score. They use a single high interest score and disregard the other scores. When we consider all of the interest scores for an individual, we have a profile of scores or a pattern of interests. The meaning of interest patterns is not a settled matter. Counselors in using the Strong Vocational Interest Blank have found that certain patterns of interest frequently occur. Further, they have found that these profiles are similar to those of persons in certain occupations. For example, Strong found that of 22 varieties of public administrators, 19 had high interest scores in "personnel manager" on his test[3]. When their profiles were compared with personnel men in industry, it was found that two groups could be differentiated on the basis of their supporting interests. The men from industry had such other interests as "production manager, office worker, accountant, purchasing agent, president, and sales manager." The public administrator's interests, on the other hand, were not as closely related to business. Their supporting interests were "lawyer, city school superintendent, social science teaching," and the exception, "production manager." If then, we were to counsel with a youth rating high in "personnel manager," much valuable information could be gleaned by examining the entire profile for a pattern of interests. On the basis of the supporting scores a distinction can be made between personnel manager in public administration as opposed to a similar position in industry.

INTEREST PATTERNS ARE MEANINGFUL

These patterns of interests are born out and in part explained by correlation coefficients between scales. We expect persons engaged in uplift occupations to have similar interests. The similarity between the interests of the minister, social worker, and YMCA secretary is revealed when coefficients or correlation among the scales are obtained. The pioneer work with interest patterns has been done with the Strong test. More recent evidence has been obtained by Kuder which supports the belief that interest pattern interpretation is also possible with his test. For the time being, we must accept on faith that it may be true of other tests. Certainly, we shall want to consider these patterns as we help our pupils interpret their interest tests.

Most of us believe that we could mark an interest test so that we get the kind of scores that we want. Undoubtedly, many pupils could do the same thing. But if they have adequate preparation before taking the interest test, we do not have to worry about their falsifying the test.

Without being aware of it, however, pupils' present choice of an occupation affects their scores. The choice leads them to mark activities which are in harmony with it; consequently, the score is higher in the chosen area. What does this mean to the counselor faced with the reality of counseling? Suppose we are to counsel a pupil who has high interest test scores in the mechanical and clerical areas, and claimed interest in bookkeeping. Knowing that a claimed choice tends to raise an interest score, it would be well to draw his attention to the mechanical score. Even though he was unaware of his mechanical interests, they were strong enough to be revealed by the test. At first, we might expect that the pupil will discount the mechanical interest just as we discount the clerical interest. If we are able to stimulate him to re-evaluate his choices, frequently we find that the differing interest gradually becomes dominant. This phenomenon of pupils warming up to an interest first revealed by a test is commonplace. It is one of the most beneficial outcomes of an interest testing program.

EXPRESSED INTEREST AFFECTS SCORES

The basic interests seem to be relatively stable from an early age. But the expression of these basic interests may take a variety of different forms during the pupil's growth. As the pupil becomes more mature, expressions of interests tend to stabilize. Before this time, however, interest tests have limited occupational implications.

TESTS OF SPECIAL APTITUDE

In the section dealing with selection of tests, we discussed special aptitude tests. At that time, we reached the conclusion that they had several limitations. Let us now see how these limitations affect their use in counseling.

1. *Special apitude tests are better bases for "No" than "Yes."* Consider the work of a watchmaker. To be successful, he must be able to comprehend things mechanical. He must have a steady hand. He must demonstrate unusual dexterity in working with small objects and tools. His eyesight should be good. We could continue listing the aptitudes that he needs for success. They are numerous. Now let us try to help Ned decide whether or not he should attempt training as a watchmaker. We get Ned to consider his interests, opportunities for training, and available evidences of ability. During the counseling process he raises the question of whether or not there are tests that will help him decide. We select two special aptitude tests: mechanical comprehension and tweezer dexterity. On the comprehension test he ranks in the upper 10 percent of high-school boys. The tweezer dexterity results place him at the fifth percentile of high-school boys. What implications do these tests have for him?

Can we say that because he is in the upper 10 percent on the mechanical comprehension test that he will be successful at watchmaking? The obvious

[3]E. K. Strong, *Vocational Interests of Men and Women* (Stanford University, California: Stanford University Press, 1943), p. 437.

answer is "No." Just because he has this aptitude, we have no assurance that he has the other necessary aptitudes. About all we can say is: "One of the aptitudes necessary for success in watchmaking is mechanical comprehension. On the basis of this test, it would seem that you have this aptitude. This does not mean you will be a successful watchmaker; it means only that you meet one of many requirements." This is not a very positive answer, but it is about as far as we can go.

From the tweezer dexterity score, we can draw a more definite conclusion. Our interpretation to Ned can be worded this way: "One of the factors necessary for success in watchmaking is the ability to work with small tools and objects. The tweezer dexterity test, which you took, roughly measures this ability. Your performance on this was not as good as we expect of a potential watchmaker. In fact, about 9 out of 10 high-school boys make higher scores on this test than you did. This does not mean that you cannot learn to be a watchmaker. It simply indicates that you do not work as accurately and as quickly with small tools as others do. Because these factors would materially affect your workmanship and, in turn, your earning power, the score on this test is a strong warning to you. Perhaps you would do well to consider some other occupations where this kind of dexterity is not so essential."

To summarize, it is easier to predict failure than success. Success requires at least a minimum of wide variety of aptitudes. Even the presence of all these aptitudes can be offset by lack of motivation, social pressures, or a host of other factors. Probable failure can be predicted if we can isolate a single major deficiency. This same principle is true for tests of scholastic aptitude also. A genius may or may not succeed in college, but it is almost certain that a moron never will.

2. *Differences between scores may not be meaningful.* When counseling a pupil who has taken two or more special aptitude tests, we are interested in comparing the results of all the tests. A pupil may have a percentile rank of 75 on an art test and 45 on a mechanical aptitude test. At first glance, it would appear that the better aptitude is for art. The norms for the art test were based on all the pupils in an unselected group of high schools. We can conclude that the competition in this group is not very stiff; thus we discount to some extent the high rank. On the other hand, the norm group for the mechanical test was composed of graduates of the machine shop in the vocational school. More reliance can be placed on this score as an indicator of real aptitude. The nearer the norm group resembles the group with which the pupil will compete, the more meaningful the percentile rank.

How to interpret norm scores when the standardization group differs in a significant way from the pupils being tested is a difficult matter to

decide. Certainly the effective counselor will become thoroughly familiar with the norm group for each test he interprets. He will realize the difficulties in making differential predictions when norm groups are not comparable. He will not always accept the difference between percentile ranks or standard scores at the face value.

We should be aware of another difficulty which arises when we attempt to compare a pupil's performance on two different tests. Even though the norms for both tests are based on the same standardization group, these differences must be interpreted cautiously. Let us see why. We know that neither test is perfectly reliable. The difference between the test scores will be affected by the reliability of both scores. Even if the tests are uncorrelated the reliability of difference scores will not be greater than the average of the two test reliabilities. Ordinarily, because correlation exists between our tests, it will be significantly less reliable. We must remember then that *the differences between scores are apt to be less reliable than the scores themselves.*

3. *Special aptitude tests are not necessary for all pupils.* In our discussion of the selection of special aptitude tests, we found that group administration could usually be justified in only the mechanical and clerical fields, and then only if training facilities for these occupations are available to pupils. It is difficult to justify group use of special aptitude tests on grounds other than administrative convenience or as an entrance hurdle. Pupils in such cases are not treated as individuals. Many of the students lack interest in these fields; others are not good risks for reasons known before testing; some students will see no point to the testing and fail to take the test seriously. In these three instances the test results can have little meaning to the pupils.

Special aptitude tests should be given only when a definite need is felt for the additional data they provide. Usually testing of this nature should be concurrent with rather than precede counseling. In using such tests, many counselors are guided by the following rules:

a. Avoid using special aptitude tests with immature pupils.
b. Let the pupil know the reason for special testing.
c. Help him to understand the purpose of special aptitude tests.
d. Wait until he feels the need of evidence from special aptitude tests before testing.
e. Try to keep the pupil from putting too much emphasis on test results.
f. Prepare him for the possibility of bad news if the test does not support his tentative choices.

USING TEST RESULTS FOR ADMINISTRATIVE PURPOSES

It is difficult to say when a test is used for guidance purposes and

when it is used for administrative purposes. Both functions should have the common goal of helping Jack or Jill make an optimum adjustment. Three administrative purposes are briefly discussed.

Modern education is planned to meet the needs of pupils. Before an adequate curriculum can be planned, it is essential to know these needs.

CURRICULUM PLANNING

Schools have found it expedient to use tests as a means of discovering the present status of pupils. Usually the results of tests given for guidance purposes are summarized. These summary statistics give the administration a picture of the student body. Three important kinds of information are available if the testing program is well rounded. They reveal the level and pattern of: (1) ability, (2) interests, and (3) achievement.

Test results help teachers do a better job of teaching. An in-service training program should include the following elements:

IN-SERVICE TRAINING OF TEACHERS

1. *Teacher recognition of individual differences.* As a result of this recognition, we can expect individualized instruction within the class to meet pupil needs.

2. *Interpretation of test results.* We do not expect all teachers to be best technicians. But to be successful, they must know the child. One of the best sources of information is the cumulative record. Tests are one kind of evidence they will find in this record. To place too much emphasis on test results is just as serious as placing too little. To get the most value from the record, teachers must know how to interpret tests in their proper perspective.

3. *The relationship of tests to marking practices.* An in-service program should caution against the tendency to type a pupil as dull or bright and to assign marks on the basis of type.

USING TEST RESULTS TO ASSIST NONSCHOOL PERSONS AND AGENCIES

A counselor once quipped, "Parents need counseling more than kids." Undoubtedly, he had just finished a session with a doting mama.

PARENT EDUCATION

Test results have been found useful in discussing pupils with their parents in these ways:

1. To help parents recognize the strengths and weaknesses of their child.
2. To urge parents to utilize this knowledge in encouraging their child toward realizable objectives.
3. To influence parents to withdraw pressure toward unrealizable objectives.
4. To help parents recognize the value of parent-school cooperation in knowing the child.

One function of the guidance program is to assist pupils to take the next step. Placement is the term used to describe this process. Test results frequently can be used to assist pupils make satisfactory adjustments in the next step. It makes no difference if the tests reveal weaknesses or strengths. The object is to help the pupil. This help can best be given when a true picture is presented. If, however, we suspect that the results are not going to be used by a person familiar with tests, we are justified in furnishing only an interpretative statement of test results. This also holds true for schools or colleges to which the pupil seeks admission. A word of caution seems appropriate. Test results should be fully described so that no possible chance for confusion exists. It is not enough to send "Otis—94." There are several forms of Otis tests. The person at the other end wants to know the date of testing, complete name and form of test, raw and norm scores, rank and type of norms. As we have already noted, this information will be found on good cumulative records.

PLACEMENT

Chapter VI

IMPROVING OUR COUNSELING SKILL

THE APPLICATION of test results to the problem of the individual from his own point of view—that of evidence he must review in arriving at his conclusions—usually takes place in the counseling interview. This fact warrants a brief discussion of the kind of counseling which will take full advantage of the testing program.

HOW SKILLFUL IS OUR COUNSELING?

In the interview we come again to a place where an important decision must be made by us. Are we capable of handling the problem, or is it beyond our counseling skill? If it is a more difficult problem than we ordinarily tackle, is there a person to whom we can refer the pupil? If no referral can be made, should we try our hand at the problem? Let us take these questions one at a time.

The only way we can develop our counseling skill is by study and by counseling. The two are inseparable. If we always refer cases that require new skills on our part, it is doubtful that we can increase our skill as a counselor. Where to draw the line between those problems we handle and those we refer is difficult to decide. Many counselors make it a practice to conduct exploratory interviews and on the basis of them decide whether or not to continue with the case.

Jacob was an unusually quiet boy in class. He rarely recited unless called upon. He seemed to have few friends. Jacob's counselor, when he plotted the scattergram illustrated in Chapter V, found him in the lower right quadrant. The counselor obtained little information about Jacob's background from the cumulative record. Few anecdotes were written by his teachers. This is a common occurrence for the quiet boys and girls. The counselor checked Jacob's standing on the achievement and ability rankings, his attendance record, his outside school activities, and talked with his teachers, but did not discover any causes for the discrepancy between his ability and achievement. He was then faced with the question of going ahead with the treatment or referring Jacob to a more skilled counselor. The counselor decided to have an exploratory interview with Jacob. As he talked with Jacob, he asked the question, "Can you study at

home?" Jacob blurted out his reply, "With my old man, you can't do anything." The counselor recognized that the reply was highly charged with emotion. Handling this kind of problem was beyond his skill. The counselor closed the interview in a manner that left Jacob at ease. After the interview, he began investigation and found that the home situation was extremely poor. The father was unemployed. He had drawn all his unemployment compensation and was living on relief. He seldom held a steady job. A logical conclusion for the counselor to make was that the father's character and actions were disturbing Jacob's personal adjustment which in turn influenced his achievement. How was the problem solved? The counselor made contact with the county welfare department. The welfare department became interested in the boy. They paid for a short series of psychiatric interviews which enabled Jacob to make a satisfactory adjustment. Not all cases will work this easily. But the point to be remembered is that an exploratory interview can be used as a judgment-making device to decide who will help the pupil. If the exploratory interview records a problem which we can handle, we go ahead.

The following suggestions have helped some counselors do a better job during the exploratory interview:

FOURTEEN SUGGESTIONS FOR COUNSELORS

1. Get ready for the interview by studying all available data carefully.
2. Prepare a plan and purpose for each interview but do not hold to it rigidly if the pupil brings other problems.
3. Get the pupil to talk; do not try to tell him.
4. Put the pupil at ease during the interview but do not let him be so much at ease that he does not think.
5. Try to interview as though you were the pupil's equal, do not give him cause to think of you as a critic or judge.
6. Admit that you do not know the answer to a question; do not bluff.
7. Be interested in what the pupil says, but do not be so interested that you try to write it all down during the interview.
8. Ask questions which cannot be answered with yes or no, but do not make them so difficult that the pupil cannot understand them.
9. Try to keep the conversation from stopping, but do not be afraid of a pause while the pupil thinks.
10. Be alert for leads which can be followed, particularly those of personal adjustment.
11. Do not express values on what the pupil says. Disgust, astonishment, or indignation have no place in the interview.

12. Have a positive suggestion to leave with the pupil or a definite date for the next interview.

13. End the interview as soon as you cease to make progress, do not let it fall to the level of inconsequential conversation.

14. Get the pupil to summarize the interview; do not let him leave with a group of ideas which do not appear related to him.

In addition to a solution of problems by means of interviews, the counselor has other tricks up his sleeve. One is seeking a solution by changing the environment. Suppose that Alex and his teacher are at swords' points. By deft use of interview techniques, we may bring a satisfactory adjustment between the two. But if the exploratory interview indicates that the process is going to take a long time, a short cut may be to suggest to the person in charge of scheduling the desirability of changing Alex's teacher. Normally, the counselor should, of course, neither desire nor be given the authority required for direct action, since his relationship with his fellow teachers and with pupils should be kept advisory and cooperative rather than administrative.

This adjustment of schedules to avoid clashes is not a weak way out, although some educators would have us think so. Stop a minute and remember all the teachers we have had. Did we like some of them as well as a few or even most of our classmates? Could we learn equally well from all of them? Hardly! If one idea is basic to the guidance movement, it is that individuals differ and have unique personalities. To expect that we shall have no clashes between teachers and pupils denies the existence of individual differences in the area of personal relations. It is not desirable to try to fix blame for these clashes. Even the best of teachers for no apparent reason will occasionally fail to establish good relations with a pupil. The wise counselor will be on the lookout for these clashes and do all he can to alleviate the situation. He should, without fail, plan some aid for any pupil whose personal adjustment is being upset by such a clash. To summarize, pupil-teacher clashes are to be expected. Where feasible, it is a good guidance technique to facilitate the rearrangement of the pupil's schedule so that he has a different teacher.

Now let us pass on to our second question: "If it is a more difficult problem than we ordinarily tackle, is there a person to whom we can refer the pupil?" Before an effective referral can be made, we must be familiar with the nature of the problem. First of all, to do this we use the exploratory interview in which we try to size up the situation. Then to what places can we refer the student for help? The following list is not inclusive, but it is suggestive of the sources of help with which the counselor must be familiar. Within the school other counselors, teachers, or deans may have a specialized skill for counseling pupils with certain problems. In the community the following persons or agencies frequently accept referrals for certain types of services: American Red Cross; Council of Social Agencies or the Community Chest; Department of Public Welfare; Department of Public Health; physicians and psychiatrists; service clubs such as Rotary, Lions, Elks, or Kiwanis; State employment services: and the counseling centers of local universities or colleges.

HANDLING DIFFICULT PROBLEMS

At times we are faced with a problem so perplexing that we doubt our ability to handle it successfully. We are in a position similar to a person who arrives at the scene of an accident. Even if he does not have medical training, he cannot escape the necessity for doing as much as he can. He renders first aid. He does those things which he can do safely. He calls for help.

Let us return to the case of Jacob. He is the underachiever who had trouble with his father. The counselor had enabled him to receive psychiatric treatment through the welfare department. Assume that Jacob's counselor was unable to arrange for psychiatric treatment. Should he tackle the problem and do the best he can? An answer of "yes" is fraught with danger. The counselor's well-intentioned actions, just as those of other staff members, are sometimes misconstrued by parents, pupils, or other citizens of the community. This danger is inherent in many of the problems with which the counselor must deal. If, for example, the counselor had asked Jacob, "What's wrong with your father?" a full-blown discussion of his father's characteristics would probably have followed. Some time later when Jacob was having an argument with his father, he may well have blurted out, "Even the counselor asked what's wrong with you!" We can visualize the counselor being summoned to the principal's office to explain his words to an angry father the next morning.

What, then, is the answer? The easiest and safest way might be to avoid the problem. Unfortunately this is not entirely feasible. We have a responsibility to the pupil just as the person at the accident has a responsibility to the injured. We must render first aid remembering the limitations of our lack of training and skill. We must continue our search for any immediate means of helping the pupil make a satisfactory adjustment, and for more remote means dependent upon our professional growth.

SHOULD PUPILS KNOW THEIR TEST RESULTS?

We come now to a controversial point. Both sides can support their arguments with equally strong emotional zeal. Should pupils know the results of tests they have taken? It is just as impossible to answer yes as no. In some cases, the answer is Yes with a capital Y and in others it is just as emphatically No. The situation is not hopeless, however.

Six criteria are discussed below. We can use them to decide when, how much, and what kind of information to give the pupil. Although this bulletin deals only with information from testing, these same criteria can be applied equally well to data secured from other sources.

1. *Only the counselor should supply pupils with test results.* The counselor is the key man in this supply line. He regulates the flow of information. In some schools, other members of the staff may reveal data during their informal counseling contacts with students. They should limit themselves to the type of information for which they have training and skill to interpret. The practice of having all members of the staff give out test scores regardless of their ability to do so is unethical. As we continue our discussion of these criteria, we shall see that decision about the nature of information to be given must be made. Making these decisions requires all the skill and training that a counselor can master.

2. *Information should be precise.* Certainly we do not want to give pupils the raw score on tests. Some pupils can understand the concept of a percentile rank within a norm group. The Bixlers believe that we should give the pupil "simple statistical predictions based upon the test data."[4] The pupil should be encouraged to apply these predictions to himself. An example of this kind of prediction is:

> A student in the upper 10 percent of his high-school class and in the upper 25 percent on a college aptitude test might be told, "We found that the best indication of success in most college courses is how well you do in high school and how you rate on a learning ability test. You were in the upper 10 percent of your high-school class and exceeded 7 or 8 out of 10 college students on a learning ability test. Most people with scores like that learn complex things relatively easily and quickly. For example, most students with scores like yours would succeed in college and get better than average grades." The last sentence of the interpretation then might be "Eighty out of 100 students with scores like yours would succeed in college and 60 would get better than average grades."[5]

But, do we give this type of information to all pupils? No. Our decision is based in part upon the next criterion.

3. *Information should be given only if the pupil can interpret it.* How do we know that the pupil can interpret it? The answer must be given by the counselor. Some counseling experts refer to the counselor's clinical judgment. By this they mean the skill that he has developed as the result of his training and experience. Ordinarily counselors base this judgment on a large number of factors just as a doctor makes his diagnosis by considering many symptoms.

4. *Pupils should be ready for test results.* If a pupil is upset emotionally because he is failing, he is not prepared to accept test data. If he has been given a test and does not know what it measures, he is not ready for the results. If the pupil has a blind faith in test results, regarding them as almost magical instruments, we had better set him straight on the value of tests before we report results. If extreme pressures are operating on him, such as parental influence or desire for approval of the group, he is not ready to accept test results unless they harmonize with the pressures that motivate him. If he is not ready for test results, they will be of little value to him.

5. *Pupils should be willing to use test results.* Most pupils ask, "How did I do on the exam?" Counselors should not accept this natural curiosity as evidence that the pupil is willing to accept test results. Unless a student is willing to use the results, to consider their implications as he makes decisions, it is unwise for the counselor to spend time discussing them. Many of the factors which influence the pupil's willingness are related to his emotional stability and maturity. Such conditions as feelings of inferiority, worries over sexual adjustment, or overcompensation for physical or mental limitations cause pupils to reject any evidence, including test data, which is not in harmony with their view of themselves. Getting information which they are not willing to accept may actually increase their emotional maladjustment. Prejudices also deter pupils from dealing objectively with information about themselves. The white-collar complex, the drive to get rich quick, and inaccurate or incomplete information cause pupils to reject test results. The counselor should be sure, therefore, that the pupil has a genuine desire to use the test results before providing him with information.

6. *The pupil should be able to take action on the information the tests give him.* Two examples will illustrate this criterion. A twelve-year-old boy in the sixth grade ranks at the first percentile of sixth grade pupils in Minnesota on two different scholastic aptitude tests. There is no point in supplying him with this information. The law requires him to remain in school. There is not much hope that he will ever do well in school even under pressure. He has no choice. There is no decision he can make which would be influenced by knowledge of his limited scholastic aptitude. This does not mean, however, that the counselor and teacher should sit idly by. They can take action on the test results and provide him with meaningful educative experience.

In the same grade, a boy with average scholastic aptitude is found to rank at the fifth percentile on an achievement test in arithmetic. Examina-

[4] R. H. and V. H. Bixler, "Test Interpretation in Vocational Counseling," *Educational and Psychological Measurement*, VI, No. 1 (1946), 145-46.
[5] *Ibid.*, 149.

tion of the test reveals that his chief difficulty is with long division. On this information, he can take action. He can remedy his deficiency.

Should pupils know their test results? The question can be answered by combining our six criteria to form this rule: *The counselor should supply pupils with as precise information as they can interpret, and on which they are ready, willing, and able to take action.*

TESTING AS A TOPIC FOR GROUP DISCUSSION

Group discussion of testing can be used to facilitate both testing and counseling. Three ways in which group discussion does this will be described.

As we counsel with pupils, it becomes apparent that they do not understand the basic concept of individual differences. They find it hard to believe that they cannot do everything equally well. Most pupils are aware of differences in scholastic aptitude, but somehow they forget that they fit into the picture, too. It is hard for all of us to accept our limitations. If pupils have an opportunity to discuss test results, the counseling process has a base from which to start. It will not have to be interrupted to provide for instruction or setting the stage, before test results are considered in the interview. An effective means of handling these discussions is to present a summary of test results given in the school. A distribution of scores might be placed on the blackboard, and through group discussion, the following points developed:

ASSISTS PUPILS UNDERSTAND INDIVIDUAL DIFFERENCES

1. *Scores have a wide range which reflects differences among individuals.* Questions such as these can be used to stimulate discussion: What is the difference between the highest and lowest score? How do you account for this difference? Would we get differences as large as this on another test? On measures of height or other physical traits? Can you think of any human characteristic where there are no differences among individuals?

2. *Most scores are found in average group.* The following questions are suggested. What 10 scores do most pupils get? How do you account for this bunching of scores? Does this make the extremely high score more significant than the average scores?

3. *Individuals may have high scores in one test and low in another.* Why do not pupils get the same marks in all subjects? Is it because pupils have more ability along some lines?

Ideally, testing of the individual pupil should follow the preliminary interview which we considered early in the chapter. The tests should be selected to meet the needs of each student. If this procedure is not possible, we accept group testing as a compromise. The advantage of this method over no testing is its only justification. One of the most perplexing problems in group testing is the attitude of pupils toward testing.

The variety of undesirable attitudes fall in three categories. They are: (1) "I don't care." These pupils approach the testing as though it were a lark. Their scores are usually too low. (2) "Makes me nervous." Emotional control is lost; the pupils become tense and rigid. This decreases their efficiency. (3) "What is it all about?" The hustle, the interruption of the normal schedule, the secrecy of test results all contribute to the feeling of confusion. The pupil's mind is filled with a variety of fears. He comes to the testing situation with countless stray thoughts running through his head.

ESTABLISH RAPPORT PRIOR TO GROUP TESTING

Each of these attitudes can be traced directly to lack of understanding. A group discussion of testing will help pupils gain a proper perspective. These discussions should stress the facts that: (a) testing is designed to assist the pupil; (b) punitive action will not result from test performance; (c) test results will be discussed with pupils during counseling, and (d) tests are only one of the judgment-making devices that teachers, counselors, and pupils use to make plans. If these discussions are successful, such problems as cheating or fake reasons for absence on the day of testing will be reduced.

ESTABLISH RAPPORT FOR COUNSELING

The preparation of pupils for counseling also will save time. If we discuss tests with a group of pupils before counseling, the first step is taken toward good rapport. The pupil knows us, he has seen us before, and he has probably come to the conclusion that we are "OK." The process of breaking the ice is a crucial point in counseling. Any effort we make toward an interesting presentation to the group will be repaid in time saved and rapport established during subsequent counseling.

Appendix A

A BASIC LIBRARY ON TESTING

BEFORE WE BEGIN the testing program in our school, we may desire much more information than is provided by this book. Certainly, as our program develops, we shall have need for reference books. In making our selection of books, we should include at least one of each of these types: (1) a basic discussion of testing and underlying theories; (2) a bibliography and critical review of tests; and (3) an elementary statistics text. In addition we should have two or more books which deal with (4) the use of tests in the guidance program. This will provide a well-rounded basic library on testing. The following six books were selected with these principles in mind. Their total cost is about $20. Books which are similar in content would be just as useful. The main objective should be to get as complete and balanced a collection as possible.

1. Bingham, Walter V. *Aptitudes and Aptitude Testing.* New York: Harper & Bros., 1937. Pp. 390. $3.00.
*2. Buros, Oscar K. *The Nineteen Forty Mental Measurements Yearbook.* Highland Park, N. J.: The Gryphon Press, 1941. Pp. 674.. $6.00.
3. Darley, John G. *Testing and Counseling in the High-School Guidance Program.* Chicago: Science Research Associates, 1943. Pp. 224. $2.95.
4. Germane, Charles E. and Edith G. *Personnel Work in High School.* New York: Silver Burdett Co., 1941. Pp. 599. $4.00.
5. Guilford, Joy P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill Co., 1942. Pp. 333. $3.35.
6. Traxler, Arthur E. *Techniques of Guidance.* New York: Harper & Bros., 1945. Pp. 394. $3.50.

Each of the six books listed above deals with certain topics which are not found in the others. On some topics most of the books have at least a short discussion. One book discusses a phase of some problem while another treats some other aspect of the problem. Thus, it is unlikely that we shall get a well-rounded discussion of testing by study of a single book. Rather, we shall have to read several volumes. The topical index below was prepared to help in finding additional material in these six volumes on the topics discussed in this book.

*Earlier editions of this book available from Rutgers University Press, New Brunswick. N. J.

Appendix B

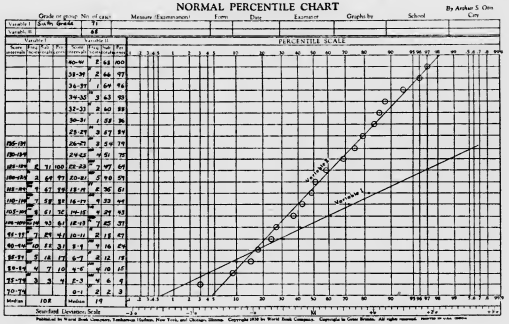
HOW TO COMPUTE LOCAL NORMS

COMPUTING PERCENTILE NORMS

Of the several procedures by which percentile norms can be computed, the graphic method is the simplest. The minor inaccuracies involved in this method are ordinarily insignificant in comparison with the errors of measurement in the raw scores themselves.

Although the percentile graph may be constructed on ordinary graph paper, use of the Otis *Normal Percentile Chart*[1] considerably simplifies the work. Therefore, the Otis Chart will be used to illustrate the procedure by which the raw scores of a local norm group can be converted to percentile ranks. There are seven steps in this process: (1) determining the score intervals; (2) tallying the frequencies; (3) finding the sub-total for each score interval; (4) computing the percents for each score interval; (5) locating points on the graph representing these percents; (6) drawing a line through these points; and (7) constructing a conversion table based on the graph.

CHART I



[1]Published by the World Book Company, Yonkers-on-Hudson, N. Y. This copyrighted chart is reproduced as Chart I by special arrangement with the publisher.

1. *Determining the score intervals.* The heavy horizontal lines on the chart divide it into 21 bands or score intervals. The range of scores to be included in each band must be such that the highest and lowest scores on the test fall within the limits of the chart. In the interest of accuracy, there should be at least 9 score intervals on the chart. There is a simple procedure for determining the score intervals which will meet these requirements. First, scan the scores to be tabulated to determine the lowest and highest scores. Subtract the former from the latter. If the difference between these two scores is:

a. 20 or less, use a score interval of 1;
b. 21 to 41, use a score interval of 2;
c. 42 to 104, use a score interval of 5;
d. 105 to 209, use a score interval of 10.

This may be called the score-interval rule.

In the first four columns of Chart I, under Variable I, are consolidated the data from a norm group of 71 sixth-grade pupils. The lowest score was 76 and the highest, 128. The difference of 52 between these two scores indicates a score interval of 5, according to the score-interval rule.

Work with the chart will be simpler if each score interval begins with a multiple of 5. Accordingly, instead of beginning the first interval with the lowest score, 76, it is begun with 75 to 79. The second interval includes scores 80-84; the third, 85-89; and so on. It will be noted that although no scores below 75 were found in the norm group, the score interval 70-74, at the lower end of the chart, and the intervals 130-134 and 135-139, at the upper end of the chart, have been included so that norms can be computed for these scores. Having determined the score intervals to use, step 2 is next.

2. *Tallying the frequencies.* This process consists simply of putting a tally mark in the second column of the chart opposite the appropriate score interval for each of 71 scores. The first score is 116 so a tally is made opposite the score interval 115-119; the second is 90, so a tally is made in the 90-94 band; the third is 109, so a tally is made in the 105-109 row, and so on until all the scores have been tallied. The tally marks are thus added and the total put in the lower right-hand corner of each block. It is well at this point to add this column of figures to make sure that the total agrees with the number in the norm group. If the sum is correct, step 3 is next.

3. *Finding the sub-total for each score interval.* The third, or sub-totals, column of the chart is obtained by cumulatively adding from bottom to top the frequencies tallied in column two. Referring to Chart I again, there were only 3 scores of 79 or less, so a 3 is entered in the sub-total column for the lowest score interval. By adding the 4 scores in the 80-84 score interval, a total of 7 scores of 84 or less is obtained as the next sub-total entry. The 5 scores in the 85-89 score interval give a sub-total of 12 scores of 89 or less. This process is continued until the highest score interval is reached in which any scores are recorded. Since this last entry tells how many scores were equal to or less than the highest score, obviously it should be the same as the total number of pupils in the norm group. As the final entry is 71, and as there are 71 pupils in the group, it is safe to go on to step 4.

4. *Computing the percents for each score interval.* To complete this column, it will be necessary to compute what percent of our entire group made scores equal to or less than the highest score in each score interval. In the example, it is clear that 100 percent of the pupils made scores equal to or less than 129 since there were no tallies above this score interval. Of course, the same result would be arrived at if the sub-total entry of 71 for this score interval were divided by the total number of pupils in our group and the quotient multiplied by 100. How should the entries for the rest of the percent column be computed? The sub-total entry for the 120-124 score interval shows that 69 pupils have scores of 124 or less. Dividing 69 by the total group of 71 gives 97 percent having scores of 124 or less. Again, dividing 67 by 71 reveals that 94 percent of the total group have scores of 119 or less. This process of dividing each sub-total by the total number in the group is continued until all entries have been made in the percents column. Now for step 5.

5. *Locating points on the graph representing these percents.* To plot the points on the graph which are represented by these percents the heavy horizontal line *above* each entry in the percent column is followed across the graph until it intersects the vertical line representing the percent entry. To illustrate, 4 percent of the pupils made scores of 79 or less. The heavy horizontal line immediately above the 4 across the graph is followed to the point where it intersects the vertical percentile line labeled 4 at the top and bottom of the chart. A small, distinct dot is made at this intersection. In the same manner for the next group, the heavy line above the 10 in the percent column is followed until it intersects the vertical percentile 10 line and a second dot is made. This process is repeated until all the percent entries have been plotted. The 100-percent point cannot be located since there is no vertical percentile line for 100. With the points plotted, step 6 is next.

6. *Drawing a line through these points.* One advantage of the *Normal Percentile Chart* over ordinary graph paper is that, in many cases, the points located will lie approximately along a straight line. If there is an abnormally large number of high or low scores, a freehand curve may have to be drawn to pass near or through the points. But in most instances,

straight-edge can be adjusted to the dots so that the line will pass through some of them and as near as possible to all of them. Draw a very light line first and then adjust it if necessary. The dots missed above the line should approximately balance the dots missed below the line. In the sample, a line was drawn through three of the dots. Three dots above the line and four dots below were missed by small amounts. If the points cluster reasonably well about this line, the final process, step 7 is next.

7. *Constructing a conversion table based on the graph.* Fine lines are drawn horizontally across the graph portion of Chart I so that each score-interval band is divided into five parts. These fine lines will help a great deal when there is a score interval of 5 and are of considerable assistance when a score interval of 2 or 10 is used. If the score interval is 1, they may be ignored. Returning to the illustration, the heavy line immediately *below* the 125-129 score interval represents a score of 125. The first fine line above this heavy line then will represent a score of 126; the second, a score of 127; the third, a score of 128; and the fourth, a score of 129. The next line on the graph is a heavy one and, since it is immediately *below* the 130-134 score interval, it represents a score of 130. The elaborate discussions of statisticians which demonstrate that a score of 125 probably should be located mid-way between the 125 and 126 line need not be of concern. In the practical situation, the scores are not sufficiently reliable to justify such refinements in procedure.

The purpose is to prepare a conversion table which will show what percentile rank should be assigned to each score on the test. Such a table, based on the graph of the Variable I data, is presented in Table 1. How did the graph permit the construction of this table? First, note that the vertical 99 percentile line cuts across the line drawn on the graph at a point which represents a score of 131. Scores of 132 or more, therefore, may be given a percentile rank of 99+. Our first entry in Table 1 is made. Since a vertical percentile line of 98.5 would intersect the line drawn at the point representing a score of 129, scores between 129 and 131 should be given a percentile rank of 99. This is the second entry in Table 1. The 97.5 percentile would intersect the line drawn well above the point representing a score of 126, so that the ninety-seventh percentile includes scores of 127 and 128. We proceed in this manner until we come to ninety-first percentile line. There is a problem. A percentile value of 92 has already been assigned to the score, 119. On investigation, it is found that the ninetieth percentile crosses the line we have drawn at a point representing a score of 118. There is no score on the test which is given a percentile rank of 91, which indicates that it is time to modify the method of locating percentile values.

Thus far we have started with the vertical percentile lines, and have found the appropriate scores. Throughout the middle range of the table, it is more convenient to start with each horizontal score line and find the percentile value which should be assigned to it. It is easy to locate the fine horizontal line representing a score of 117 and to see that it intersects the line drawn at the 89th percentile. Continue to locate the percentile values of each score in this manner until the extremely low scores are approached. There, it is frequently necessary to return to the original process of starting with the vertical percentile lines and locating the scores which should be assigned to each percentile rank. With the conversion table complete, the standing of my pupil's score with regard to this norm group can be quickly ascertained.

TABLE 1

TABLE FOR CONVERTING VARIABLE I RAW SCORES TO PERCENTILE RANKS BASED ON DATA SHOWN IN CHART I.

Raw score	Percentile rank	Raw score	Percentile rank	Raw score	Percentile rank
132 & above	99+	110	75	93	24
129–131	99	109	72	92	22
127–128	98	108	69	91	19
125–126	97	107	66	90	17
123–124	96	106	63	89	15
122	95	105	60	88	13
121	94	104	57	87	11
120	93	103	54	86	10
119	92	102	51	85	9
118	90	101	48	84	8
117	89	100	45	83	6
116	87	99	41	81–82	5
115	85	98	38	80	4
114	83	97	35	78–79	3
113	81	96	32	75–77	2
112	79	95	29	70–74	1
111	77	94	27		

Chart I also includes data for Variable II based on the scores of 68 tenth-grade pupils in the norm group. The lowest score on the test was 1 and the highest 41. Check each of the seven steps for these data. The conversion table which is constructed from the Variable II line is presented in Table 2.

Close study of Table 1 will reveal one aspect of percentile norms which was mentioned briefly in Chapter III. If a score is either very high or very low, a one-point change in that score will change its percentile little if at all. If a score is near the middle of the table, however, a one-point change in the score will change its percentile value considerably. In other words, a difference between two percentile ranks near either end of the table is more significant than the same difference between two percentiles near the middle of the table. The point is illustrated by the spacing of the vertical percentile lines in Chart I. The space between the 1 percent and the 2 per-

TABLE 2

TABLE FOR CONVERTING VARIABLE II RAW SCORES TO PERCENTILE RANKS BASED ON DATA SHOWN IN CHART I.

Raw score	Percentile rank	Raw score	Percentile rank	Raw score	Percentile rank
41 & above	98	27	77	13	30
39–40	97	26	74	12	27
38	96	25	71	11	24
37	95	24	68	10	21
36	94	23	65	9	19
35	93	22	61	8	16
34	91	21	58	7	14
33	90	20	54	6	12
32	88	19	51	5	11
31	86	18	47	4	9
30	84	17	44	3	8
29	82	16	40	2	6
28	79	15	37	1	5
		14	33	0	4

cent lines is greater than the space between the 40 percent and the 50 percent lines. The difference between a percentile rank of 1 and a percentile rank of 2 is greater than the difference between a percentile rank of 40 and a percentile rank of 50. Although the percentile rank on a single test is relatively easy to interpret, this irregularity of the percentile scale makes comparison of ranks on several tests rather difficult. Standardized scores are not as easily understood by the layman, but they are perfectly regular in a normal distribution. For this reason, the following table for converting percentiles to standardized scores is presented.

TABLE 3

TABLE SHOWING EQUIVALENT VALUES OF PERCENTILE RANKS AND STANDARDIZED SCORES IN A NORMAL DISTRIBUTION.

Percentile rank	Standardized score	Percentile rank	Standardized score	Percentile rank	Standardized score
99+	76	75–77	57	23–25	43
99.	72–75	71–74	56	20–22	42
98.	70–71	68–70	55	17–19	41
97.	68–69	64–67	54	15–16	40
96.	67	60–63	53	13–14	39
95	66	56–59	52	11–12	38
93–94	65	52–55	51	9–10	37
92	64	49–51	50	8	36
90–91	63	45–48	49	6– 7	35
88–89	62	41–44	48	5	34
86–87	61	37–40	47	4	33
84–85	60	33–36	46	3	31–32
81–83	59	30–32	45	2	29–30
78–80	58	26–29	44	1	25–28

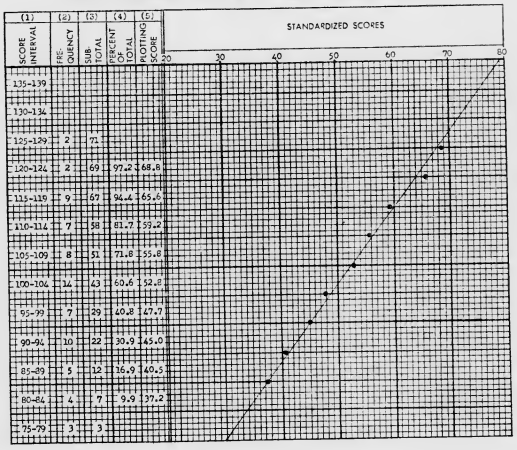
COMPUTING STANDARDIZED SCORE NORMS

Frequently only standardized scores are desired. In this case, computing percentiles and then converting them to standardized scores by use of table 3 is not economical. A graphic method can be used for computing standardized scores. There are eight steps in this method of converting raw scores of a local norm group to standardized scores. They are: (1) preparing the chart; (2) determining the score intervals; (3) tallying the frequencies; (4) finding the sub-total for each score interval; (5) computing the plotting scores; (6) locating points on the graph representing these plotting scores; (7) drawing a line through these points; and (8) constructing a conversion table based on the graph.

1. *Preparing the chart.* Ordinary graph paper may be used in preparing the chart. The layout is illustrated in Chart II. It is advisable not to draw the horizontal lines until after the score interval has been determined by step 2. But the heavy vertical lines and the headings of the columns should be made. Seven standardized scores are indicated at the top of the chart. The five columns to the left are numbered to make it easy to refer to them in this discussion.

CHART II

GRAPHIC METHOD FOR COMPUTING STANDARDIZED SCORES



2. *Determining the score intervals.*

3. *Tallying the frequencies.*

4. *Finding the sub-total for each score interval.*

These three processes are exactly the same as in computing percentile ranks with one exception. It will be necessary to draw the horizontal lines separating the score intervals after the size of each score interval (step 2) has been determined.

5. *Computing the plotting scores.* There are two steps to this process.

a. Find what percent each sub-total is of the total. The percentages for the highest and lowest intervals need not be computed. On Chart II the sub-totals are found in column 3. Begin with the second sub-total which in this case is 7. It is 9.9 percent of the total 71. This is recorded in column 4. The next sub-total, 12, is 16.9 percent of 71. It is also recorded in column 4. After each percentage has been computed, the next step may be begun.

b. From Table 4 obtain the plotting score corresponding to each of the percents recorded in column 4. The plotting scores are recorded in column 5 of Chart II. It is satisfactory to round the percentages to the nearest whole number. Thus in our example 9.9 percent is rounded to 10. The plotting score for 10 percent is 37.2. In like manner, the remaining plotting scores are obtained.

6. *Locating points on the graph representing these plotting scores.* The procedure for locating the five points on the graph which represent plotting scores is similar to step 5 of the percentile norm process. The points should be plotted on the line at the *top* of the score interval. Thus the plotting score for the 80-84 score interval is plotted on the horizontal 85 line slightly to the right of its intersection with the vertical 37 line. In like manner the remaining plotting scores are plotted.

7. *Drawing a line through these points.* This process is exactly the same as step 6 of the percentile norm procedure.

8. *Constructing a conversion table based on the graph.* This process is similar to step 7 of the percentile norm procedure. Review the first paragraph of this discussion. When it is understood how to locate the horizontal lines representing each score the method by which Table 5 was constructed may be followed.

Close examination of Chart II reveals that the top horizontal line representing a score of 139 cuts across the line drawn on the graph very close to the vertical standardized score line of 79. This is the first entry in Table 5. The horizontal score line of 138 crosses the line drawn at the vertical standardized score 78. This is the second entry in the table. The 137 score-line cuts the line drawn close to the vertical standardized score of 77—the third entry in the table. The fourth entry, 136, equals 76. Both the 135

TABLE 4

A Table for Converting Percentages to Plotting Scores for Graphic Method of Computing Standardized Scores.

Percent	Plotting score	Percent	Plotting score	Percent	Plotting score
1	26.7	34	45.9	67	54.4
2	29.5	35	46.2	68	54.7
3	31.2	36	46.4	69	55.0
4	32.5	37	46.7	70	55.3
5	33.5	38	47.0	71	55.5
6	34.5	39	47.2	72	55.8
7	35.2	40	47.5	73	56.1
8	36.0	41	47.7	74	56.4
9	36.6	42	48.0	75	56.7
10	37.2	43	48.3	76	57.1
11	37.7	44	48.5	77	57.4
12	38.3	45	48.7	78	57.7
13	38.7	46	49.0	79	58.1
14	39.2	47	49.3	80	58.4
15	39.6	48	49.5	81	58.8
16	40.1	49	49.7	82	59.2
17	40.5	50	50.0	83	59.5
18	40.8	51	50.3	84	59.9
19	41.2	52	50.5	85	60.4
20	41.6	53	50.7	86	60.8
21	41.9	54	51.0	87	61.3
22	42.3	55	51.3	88	61.7
23	42.6	56	51.5	89	62.3
24	42.9	57	51.8	90	62.8
25	43.3	58	52.0	91	63.4
26	43.6	59	52.3	92	64.0
27	43.9	60	52.5	93	64.8
28	44.2	61	52.8	94	65.6
29	44.5	62	53.0	95	66.5
30	44.7	63	53.3	96	67.5
31	45.0	64	53.6	97	68.8
32	45.4	65	53.8	98	70.5
33	45.6	66	54.1	99	73.3

and 134 horizontal score lines intersect the line drawn at points closer to standardized score, 75, than to any other standardized score. For the fifth entry in Table 5, it may be said that scores of 135 and 134 are both equivalent to a standardized score of 75. Proceed in this manner until all scores on the chart have been assigned standardized scores.

ESTIMATING ACCURACY OF STANDARDIZED SCORES

One advantage of standardized scores has not been mentioned. If the reliability of the test is known, it is possible to estimate how accurately the test is measuring in terms of standardized score units. Since most of the tests mentioned in this book have reliability coefficients around .90, this degree of accuracy has been assumed in preparing Table 6. Two examples will suffice to illustrate how this table may be used. Suppose that a pupil's true standardized score on a reading readiness test is 65. This true score is

TABLE 5

TABLE SHOWING RAW SCORES AND EQUIVALENT STANDARDIZED SCORES BASED ON DATA SHOWN IN CHART II.

Raw scores	Standardized scores	Raw scores	Standardized scores	Raw scores	Standardized scores
139	79	115–116	61	92	43
138	78	114	60	90–91	42
137	77	113	59	89	41
136	76	111–112	58	88	40
134–135	75	110	57	86–87	39
133	74	109	56	85	38
131–132	73	108	55	84	37
130	72	106–107	54	82–83	36
129	71	105	53	81	35
127–128	70	104	52	80	34
126	69	102–103	51	78–79	33
125	68	101	50	77	32
124	67	99–100	49	76	31
122–123	66	98	48	75	30
121	65	97	47	74	29
119–120	64	96	46	72–73	28
118	63	94– 95	45	71	27
117	62	93	44	70	26

the one he would make on a similar reading readiness test having perfect reliability. Since our test is not perfectly reliable ($r = .90$), the score he makes on our test will probably vary to some extent from his true score of 65. Table 6 shows how much variation to expect. From it we can say that there are 75 chances in 100 that the score he obtains on our test will be as great as 1 point above or below his true score. There are 50 chances in 100 that his obtained score will be as great as 2 points away from his true score. His chance of obtaining a score as much as 8 points greater or less than his true score is only 1 in 100.

For a second example, consider a pupil with obtained standardized scores of 51 in clerical aptitude and 57 in mechanical aptitude. If both tests have reliabilities of about .90, Table 6 may be used to help decide

TABLE 6

TABLE SHOWING CHANCES IN 100 THAT CERTAIN DEVIATIONS BETWEEN OBTAINED AND "TRUE" STANDARD SCORES WILL OCCUR.
(Reliability coefficient of test assumed to be .90.)

Deviation of obtained score from true score	Chances in 100 of deviation occurring
1	75
2	50
3	32
4	18
5	10
6	5
7	2
8	1

whether the pupil's aptitudes in these two fields are really different. What are the chances that he would have obtained these two scores even though his true scores in each test were identical? It may be assumed that this true score in each test was 54, midway between the two obtained scores. This is a 3-point deviation in each case. Table 6 shows that there are 32 chances in 100 that his obtained score in clerical aptitude would be as much as 3 points lower than his true score. Likewise, there are 32 chances in 100 that his obtained score in mechanical aptitude would be as much as 3 points higher than his true score. What are the chances that both of these events would occur simultaneously? Statisticians say that the answer to that question is found by multiplying the chances. So 32/100 x 32/100 makes 1024/10000. Crossing off two places in the top and bottom of this fraction gives roughly 10 chances in 100 that these two scores would be obtained if the pupil's aptitudes for both mechanical and clerical work were the same. Of course, these figures may be interpreted the other way to indicate that there are 90 chances in 100 that his mechanical aptitude is superior to his clerical aptitude.

INDEX

END OF
TITLE